Advances in Automatic Target Recognition (ATR) for CT-Based Object Detection Systems - Final Report

Conducted by DHS Center of Excellence,
ALERT at Northeastern University

For

U.S. Department of Homeland Security Directorate of Science and Technology, Explosives Division

This report was supported by the U.S. Department of Homeland Security under Task Order Number HSHQDC-12-J-00429.

This page intentionally left blank.

Table of Contents

This page intentionally left blank.

1 Executive Summary

1.1 Background

Five research groups developed automated target recognition algorithms (ATR[1]) for CT-based explosive detection systems (EDS). The research groups (four from academia and one from a national laboratory) were provided with images from scans of targets packed in bags and scanned on a medical CT scanner. The targets were saline, rubber and modeling clay with certain minimum masses and minimum sheet thicknesses. The targets where chosen to create detection scenarios similar to what vendors face when detecting explosives using their CT-based explosive detection systems. Bulk and sheet targets were scanned along with stream of commerce non-targets with different amounts of clutter and different types of containment. The goal was to achieve a probability of detection (PD) greater than 90% and a probability of false alarm (PFA) less than 10%. Another researcher developed automatic scoring tools and created ground truth labels for training and testing purposes.

The project was executed by a Department of Homeland Security (DHS) Center of Excellence (COE) at Northeastern University (NEU) known as *Awareness and Localization of Explosives-Related Threats* (ALERT). The project was funded by the Explosives Division (EXD) of the Science and Technology (S&T) Directorate of DHS. The project was based on recommendations from the participants at the *Advanced Development for Security Applications* (ADSA[2,3]) workshops held at ALERT in order to support DHS's objective to increase the involvement of third-parties[4] to augment the capabilities and capacities of the vendors of explosive detection equipment[5]. In particular, DHS charged ALERT with learning how to work on relevant problems using unclassified data and putting the results into the public domain so that future researchers could build on the work from this program. A program review was held at the end of the project so that the researchers could present their results to the incumbent vendors, DHS, TSA and other third-parties.

The project was designed with the following outcomes for DHS.

- The program may improve ATRs. The improved target recognition may lead to decreased minimum target mass, increased target population coverage, increased probability of detection and decreased probability of false alarm.

[1] A table of acronyms and terms used in this report can be found in Section 10.
2 Information about the ADSA workshops, including their final reports, can be found at: www.northeastern.edu/alert/transitioning-technology/strategic-studies
[3] Most of the recommendations were made at the first two ADSA workshops: ADSA01 and ADSA02.
[4] First party is TSA. Second parties are incumbent vendors of explosive detection equipment. Third parties are academia, national labs and industry other than the incumbent vendors.
[5] A table is available to DHS and TSA upon request from Laura Parker (DHS) showing involvement of industry and the government with third parties who were brought into the security field through the ADSA workshops and other projects funded at ALERT.

- The program may increase involvement of third parties via the availability of common CT datasets and tools, which will increase the work in target recognition, and the number of students who can join the workforce of the vendors and DHS.
- The program may foster collaboration between academics, national laboratory personnel and incumbent security industry vendors.
- The project will develop methods to create projects relevant to the security field using non-classified datasets and requirements so that third parties develop technologies that can be transitioned to deployed equipment.

1.2 Outcomes

The outcomes of this project were as follows.

- ATR approaches
    - Five different ATRs were created to solve the detection requirements using, in part, the following algorithms:
        - Pixel classification prior to segmentation.
        - EM (expectation and maximization) algorithm.
        - MFT (mean field theory).
        - Markov random fields.
        - Random forests.
        - Parallel segmentation.
        - Graph-based splitting.
        - Image-based metal artifact removal.
        - Shape-based classifiers.
    - Most of the ATRs exceeded the requirements for PD and PFA.
    - The researchers worked on problems similar to what the vendors face.
    - Vendors want to assess the robustness of the ATRs.
    - Vendors expressed a desire to engage the researchers to adapt their ATR algorithms to vendor-supplied data.
- Dataset and tools
    - The scans of the targets on a medical scanner may be representative of the images generated by deployed EDS equipment.
    - A publically available dataset[6] of projection data, images and meta-data together has been created for future research into reconstruction, segmentation and ATR algorithms. Meta-data characterizes the scanner geometry, x-ray optics, readout sequences and file formats.
    - The automatic scoring tools, which are also in the public domain, simplified the development, scoring and comparison of the ATR algorithms. At least one vendor is considering using the scoring tools.

[6] http://www.northeastern.edu/alert/transitioning-technology/alert-datasets/

- Process
    - Technology foraging in the medical imaging field may benefit the security field.
    - The project may support TSA's objectives as outlined in the document entitled "Transportation Security Strategic Capability Investment Plan[7]."

The above outcomes may have the following benefits for industry, DHS and TSA:

- Learned how to engage third-party performers in the research, development and transition of explosive detection equipment.
- Detection performance may be improved with ATR algorithms developed by third parties.
- Students trained in the security field may move into positions in industry, the national labs.
- The researchers in this project may continue to work with the vendors, DHS and TSA.
- The public domain dataset may be used by future researchers leading to additional positive outcomes for DHS.
- This report, together with researchers presenting and publishing their work, may increase the size of the community working on security problems. This should allow the scientific method[8] to develop improved algorithms using the datasets put into the public domain.

1.3 Recommendations for DHS

DHS should consider[9] funding additional research with third parties in the following areas:

- Continued ATR development for CT-based EDS including increasing the number of scans used, using representative scans of explosives on TSA-deployed scanners, emulating the blind testing employed by DHS and degrading the image quality of the medical scans.
- Having the national labs apply the ATR algorithms developed for this project to sensitive security information (SSI) and classified scans from TSA-deployed equipment.
- Development of advanced reconstruction and ATR algorithms for screening divested objects at the checkpoint (TSA's term: AT2), personnel screening (TSA's term: AIT) and cargo inspection equipment including application to single- and dual-energy scanners, photon counting detectors, and sparse- and multiple-view CT scanners.
- Development of combined reconstruction, segmentation and ATR algorithms.
- Development of simulation tools and simulated test sets for equipment using x-rays and millimeter waves (MMW).
- Decreasing the time required to certify and qualify equipment using automated scoring tools and application of statistical tools.
- Reduction of the computational expense of advanced image reconstruction algorithms.

---

[7] https://www.fbo.gov/index?tab=documents&tabmode=form&subtab=core&tabid=ecd4fb9891348a3a7d339640e6d906d9
[8] http://en.wikipedia.org/wiki/Scientific_method
[9] These recommendations are based on the remaining content of this report.

2 Funding

The material in this report is based upon work supported by the U.S. Department of Homeland Security under Order Number HSHQDC-12-J-00429.

3 Introduction

3.1 Background

The Department of Homeland Security (DHS) has requirements for future explosives detection systems (EDS) that include increased probability of detection and decreased probability of false alarm for a larger set of threats and with reduced minimum masses. The larger set of threats includes certain types of homemade explosives (HME). There are indications that these requirements for future EDS equipment may be difficult to achieve with the technologies presently deployed in the field. In order to resolve these issues, DHS has adopted the strategy of augmenting the capabilities and capacities of the vendors of EDS equipment with the involvement of third parties. Third parties are defined as researchers from academia and industry other than the vendors.

DHS funded ALERT to execute a project denoted the *Automated Target Recognition (ATR) Initiative*, which is also known as *Task Order Four* (TO4). The goal of this project was to involve third parties in the development of ATR algorithms that could eventually be deployed by the incumbent vendors. The work was led by the Northeastern University component of the ALERT DHS Center of Excellence (COE). The investigators for the projects were comprised of researchers both within and outside of the current group of people being supported by the COE.

This project was based on recommendations from the participants at the *Advanced Development for Security Applications* (ADSA) workshops held at ALERT in order to support DHS's objective to increase the involvement of third-parties. In particular, DHS charged ALERT with learning how to work on relevant problems using unclassified data and putting the results into the public domain. This project is the third in a series of projects to learn how to increase the involvement of third parties. In the first project, which was denoted the segmentation initiative or Task Order 1 (TO1), researchers developed segmentation algorithms that could be employed in ATR algorithms for CT-based EDS equipment[10]. In the second project, which was denoted the reconstruction initiative or Task Order 3 (TO3), researchers developed reconstruction algorithms for CT-based EDS equipment[11]. The present project is built on the results and algorithms developed under TO1 and TO3.

The project was designed with the following outcomes for DHS.

- The program may improve ATRs. The improved target recognition may lead to decreased minimum target mass, increased target population coverage, increased probability of detection and decreased probability of false alarm.

[10] The final report for TO1 can be found at:
https://myfiles.neu.edu/groups/ALERT/strategic_studies/SegmentationInitiativeFinalReport.pdf
[11] The final report for TO3 can be found at:
https://myfiles.neu.edu/groups/ALERT/strategic_studies/TO3_FinalReport.pdf

- The program may increase involvement of third parties via the availability of common CT datasets, and tools, which will increase the work in target recognition, and the number of students who can join the workforce of the vendors and DHS.
- The program may foster collaboration between academics, national laboratory personnel and incumbent security industry vendors.
- The project will develop methods to create projects relevant to the security field using non-classified datasets and requirements so that third parties develop technologies that can be transitioned to deployed equipment.

A program review was held near the end of the project so that the researchers could present their results to the security vendors, DHS, TSA and other third-parties.

3.2 Project Overview

An overview of the project is as follows.

1. Bags were packed with target materials as well as objects found in stream of commerce in airports.
2. The target materials were saline, polymer (modeling) clay and rubber.
3. Targets of any material were considered “bulk” or “sheet” form.
    a. Most rubber targets were sheets (exception: rubber rods).
    b. Most saline and clay targets were bulk (exceptions: saline placed flat in plastic bags and clay rolled out into sheets).
4. The bags were scanned on a medical CT scanner resulting in images of the bags. The scans were single energy. Projection data were also collected for use in future research projects.
5. The contents and their placement in bags were documented in an Excel spreadsheet.
6. Pictures and videos of targets, non-targets and packed bags were taken.
7. The voxels corresponding to the targets in the images were marked and stored in label images. This marking is known as ground truth (GT) label images.
8. ATRs were developed using the images and the ground truth.
9. The desirable characteristics of the ATRs were:
    a. Minimizing the probability of false alarm (PFA) for a specified probability of detection (PD).
    b. Minimal use of algorithms for specific target configurations (known as corner cases).
    c. Minimal overtraining on test data.
    d. Novelty compared to the prior art.
    e. Ability to detect targets in difficult configurations.
    f. Potential to be extended to detect additional targets.
10. Software tools were supplied to compare the results of ATRs with the ground-truth.
11. The following items were supplied to assist the development of ATRs:
    a. A sample (notional) ATR (denoted SATR) so that common functions (e.g., reading and writing images and results) did not have to be replicated by each researcher.

b. The segmentation algorithms developed for the Segmentation Initiative (Task Order 1) were available for use in this project.
c. A bibliography describing related prior art in the ATR field.

The purpose of this report is to present the following aspects of the project:

1. Methods
2. Results
3. Discussion
4. Recommendations for additional work
5. Lessons learned
6. Limitations of this project

4 Document Layout

This remainder of this report is organized as follows:

- Section 5 presents methods.
- Section 6 presents results.
- Section 7 presents a discussion of the results including findings and recommendations for future work.
- Section 8 lists the people who helped make this program a success.
- Section 9 lists the project team.
- Section 10 contains tables of acronyms and terms.
- Section 11 contains supplemental material including the presentations made by the researchers at the program review and their final reports.

Multiple terms have been used during this project for the same idea. The variations of the terms are described in Section 10. For the most part, we tried to use only one version of each term in the body of this report. However, the different versions of the terms may be used in the supplemental material found in the appendices.

5 Methods

5.1 Statement of Work and Researcher Selection

A statement of work (SOW or requirement specification) was written for the various aspects of this project [Section 11.2.2]. The SOW divided the work into a number of tasks (also known as sub-tasks). ALERT chose research groups to perform the tasks and created contractual relationships with the research groups as necessary. The breakdown of tasks and researchers is found in the following two tables.

ATR Algorithm Research		
Task	**Performer**	**Institution**
ATR Development	Synho Do	Massachusetts General Hospital
ATR Development	Jens Gregor	University of Tennessee
ATR Development	Dong Hye Ye Charlie Bouman Pengchong Jin	Purdue University
ATR Development	Jun Zhang Laura Drake Hongquan Zuo	University of Wisconsin, Milwaukee
ATR Development[12]	Philip Top Ana Paula Sales Hyojin Kim Timo Bremer Steve Azevedo Harry Martz	Lawrence Livermore National Laboratory

Support Tasks		
Task	**Performer**	**Institution**
Scoring tools, ground truth labels, sample ATR	Franco Rupcich	Self
Scanning on medical CT scanner	Doug Boyd Sam Song	Telesecurity Sciences
Dataset development – packing, scanning, documentation	Rick Moore Alyssa White	Massachusetts General Hospital and Northeastern University
Dataset validation and level of difficulty determination	Steve Skrzypkowiak[13]	DHS S&T EXD (SETA)

[12] LLNL was funded directly by DHS to support this project under the leadership of Harry Martz.

[13] Steve Skrzypkowiak was funded directly by DHS to support this project.

5.2 Dataset Creation

A requirement specification was written for the CT scanner that had to be used for this project [Section 11.3.1]. The key points of the specification were:

1. The image quality (IQ) of the scanner should be comparable to deployed EDS scanners.
2. All data (projection, images and meta-data) has to be put in the public domain[14].
3. Sufficient meta-data should be supplied so that code can be written to reconstruct images that closely match online reconstructions.

The Imatron C-300 was chosen as the scanner for the project. This is the same scanner that was used for the reconstruction initiative (TO3).

The following targets were chosen for the researchers to evaluate:

1. Saline with varied concentrations.
2. Rubber sheets with various thicknesses.
3. Modeling clay in various shapes.

The minimum mass of targets was 250 g. The minimum thickness of sheets was 1/4 inch. Sheets thicker than 3/8 inch were considered to be bulks.

An additional set of targets were also scanned to allow development of ATRs beyond the scope of this project. These targets, which were known as pseudo targets (PT), had lower mass and density than the other targets and with thinner thicknesses of sheets. The ATRs developed for this project were required meet the PD requirement for PT sheets. These PT sheets still were at least 250 g.

These materials were chosen to be a set of reproducible objects. Other clutter objects were also scanned as well as calibration objects.

A specification was written for packing and scanning bags [Section 11.3.2].

The targets were packed in bins and bags along with stream of commerce items. The bins and bags were scanned on the Imatron medical CT scanner. The images and raw data corresponding to the scans were collected. The total number of scans was 188. The total number of targets scanned was 75. The total number of pseudo targets was 75, of which 10 were pseudo target sheets. The number of non-targets scanned in these scans was 1371. Note that the targets, pseudo targets and non-targets were scanned multiple times.

A dataset was created of the following items:

1. CT cross-sectional images.

[14] Sensitive information (e.g., SSI) cannot be contained in publications using this data. A process denoted REAP is applied to all publications before they are released into the public domain.

2. Projection data (raw and corrected).
3. Meta-data including scanner geometry and file formats.
4. Maps of objects in the CT scans known as ground truth data.
5. Descriptions of the items themselves and how they were packed.
6. Pictures and videos of the targets, non-targets and packed bags.

The database was hosted at a SFTP site at Boston University. Access was granted to researchers after they signed a confidentiality disclosure agreement (CDA). The SFTP site controls individual user access and is backed up by Boston University. Researchers interested in obtaining access to the data should contact ALERT through ALERT's website[15].

A document was created with the lessons learned during building this dataset [Section 11.3.3].

5.3 ATR Development

5.3.1 Overview

The research teams developed their ATR algorithms over a period of approximately twelve months. During this period of time, the following support activities took place:

1. A kickoff meeting, via a telephone conference, was held [Section 11.2.1].
2. A technical specification for the project was provided to the researchers [Section 11.1.1].
3. Monthly status meetings, via a telephone conference, were conducted.
4. Researchers wrote monthly status and technical reports.
5. A bibliography of relevant material was created [Section 11.2.3].

5.4 ATR Requirements

The ATRs had to have the following characteristics:

1. PD ≥ 90% on targets and PT sheets.
2. PFA ≤ 10%.
3. PD and PFA were determined using the automated scoring tools with specified precision and recall with respect to ground truth labels.
4. Detection of targets with a high level of difficultly (LOD) [Section 11.1.2] was emphasized.
5. The sample ATR could be used as a starting point.
6. The ground truth labels could be used to assist in the training process.
7. The output of the ATR was a label image showing the location of suspected targets.
8. Simulated images were provided in order to simplify development of the ATRs.
9. The input to the ATR were the reconstructed images. Projection data (raw and corrected) may be used by the ATR but not reconstructed.
10. There was not to be an upper limit on mass or volume.
11. Detection should be independent of shape, size, location, orientation, clutter, and concealment.
12. The researchers were requested to:

[15] http://www.northeastern.edu/alert/transitioning-technology/alert-datasets/

   a. Separate the data into training and test sets
   b. Not over-train on the data
   c. Design their ATRs to be extensible so that additional targets can be considered in the future.

13. Containers (e.g., bottles) for liquids were not considered to be part of the target.
14. The ATRs should be different than the methods presented in the prior art.
15. There was no requirement to report the type of target (e.g., saline, modeling clay or rubber sheet).
16. Execution speed and computational requirements for ATRs were not of consideration (i.e., out of scope) for this project.

5.5 Support Tools

The following support tools, functions and data were created for this project:

1. Ground truth labels.
2. Software (scripts) to perform automated execution of an ATR on a set of images and creation of PD and PFA reports.
3. Sample ATR (SATR).
4. Excel spreadsheet for
   a. Target and non-target characteristics,
   b. Packing information,
   c. Scanning information and
   d. Pictures and videos.

See Section 11.4 for additional details on the support tools.

5.6 Program Review

A program review was held on November 6, 2014. The following is a list of the topics discussed during the program review:

1. Project overview.
2. Presentations from the five researchers.
3. Feedback from vendors.
4. Feedback from attendees.
5. Recommendations for next steps.

The presentations corresponding to these topics can be found in Section 11.5.2. Additional material related to the program review can be found in Section 11.5.

5.7 Final reports

The researcher groups described their work in final reports [Section 11.6].

5.8 Subject Matter Experts

The following people were chosen by ALERT to be subject matter experts (SME):

1. Carl Crawford, Csuptwo.
2. David Castañón, Boston University.
3. Clem Karl, Boston University.
4. Harry Martz, Lawrence Livermore National Laboratory.

The responsibilities of the SMEs were as follows:

1. Provide technical assistance to program leadership.
2. Review datasets that were used by the research groups.
3. Mentor the research groups.
4. Organize and moderate the program review.
5. Write this final report.

6 Results

The PD and PFA results from the five groups are as follows:

Researcher	PD – all targets (%)	PD – high level of difficulty (%)	PD – pseudo target sheets (%)	PFA (%)
Synho Do - Massachusetts General Hospital	94.3	94.3	100.0	8.3
Jens Gregor – University of Tennessee	88.2	88.0	70.0	20.0
Dong Hye Ye – Purdue University	95.0	94.0	90.0	8.0
Jun Zhang – University of Wisconsin, Milwaukee	89.2	87.7	80.0	9.7
Philip Top – Lawrence Livermore National Laboratory	93.6	92.7	100.0	11.9

7 Discussion

7.1 Outcomes

The outcomes of this project were as follows:

- ATR approaches.
    - Five different ATRs were created to solve the detection requirements using, in part, the following algorithms:
        - Pixel classification prior to segmentation.
        - EM (expectation and maximization) algorithm.
        - MFT (mean field theory).
        - Markov random fields.
        - Random forests.
        - Parallel segmentation.
        - Graph-based splitting.
        - Image-based metal artifact removal.
        - Shape-based classifiers.
    - Most of the ATRs exceeded the requirements for PD and PFA.
    - The researchers worked on problems similar to what the vendors face.
    - Vendors want to assess the robustness of the ATRs.
    - Vendors expressed a desire to engage the researchers to adapt their ATR algorithms to vendor-supplied data.
- Dataset and tools
    - The scans of the targets on a medical scanner may be representative of the images generated by deployed EDS equipment.
    - A publically available dataset[16] of projection data, images and meta-data together has been created for future research into reconstruction, segmentation and ATR algorithms. Meta-data characterizes the scanner geometry, x-ray optics, readout sequences and file formats.
    - The automatic scoring tools, which are also in the public domain, simplified the development, scoring, and comparison of the ATR algorithms. At least one vendor is considering using the scoring tools.
- Process
    - Technology foraging in the medical imaging field may benefit the security field.
    - The project may support TSA's objectives as outlined in the document entitled "Transportation Security Strategic Capability Investment Plan[17]."

---

[16] http://www.northeastern.edu/alert/transitioning-technology/alert-datasets/
[17] https://www.fbo.gov/index?s=opportunity&mode=form&id=bcdf2fca93b27daf4a6a399297c25dc7&tab=core&_cview=0

The above outcomes may have the following benefits for industry, DHS and TSA.

- Learned how to engage third-party performers in the research, development and transition of explosive detection equipment.
- Detection performance may be improved with ATR algorithms developed by third parties.
- Students trained in the security field may move into positions in industry, the national labs.
- The researchers in this project may continue to work with the vendors, DHS and TSA.
- The public domain dataset may be used by future researchers leading to additional positive outcomes for DHS.
- This report together with researchers presenting and publishing their work may increase the size of the community working on security problems. This should allow the scientific method[18] to develop improved algorithms using the datasets put into the public domain.

7.2 Recommendations for DHS and Future Work

DHS should consider funding additional research with third parties in the following areas:

- Continued ATR development for CT-based EDS including increasing the number of scans used, using representative scans of explosives on TSA-deployed scanners, emulating the blind testing employed by DHS and degrading the image quality of the medical scans.
- Having the national labs apply the ATR algorithms developed for this project to sensitive security information (SSI) and classified scans from TSA-deployed equipment.
- Development of advanced reconstruction and ATR algorithms for screening divested objects at the checkpoint (TSA's term: AT2), personnel screening (TSA's term: AIT) and cargo inspection equipment including application to single- and dual-energy scanners, photon counting detectors, and sparse- and multiple-view CT scanners.
- Development of combined reconstruction, segmentation and ATR algorithms.
- Development of simulation tools and simulated test sets for equipment using x-rays and millimeter waves (MMW).
- Decreasing the time required to certify and qualify equipment using automated scoring tools and application of statistical tools.
- Reduction of the computational expense of advanced image reconstruction algorithms.

7.3 Limitations and Lessons Learned

1. As known at the beginning of the project, it is not possible to determine if the ATRs were novel or improved with respect to the ATRs developed for certified explosive equipment because the researchers did not have access to the following:
    a. Scans of threats on TSA-deployed equipment.
    b. Classified requirement specifications.
    c. Ability to take certification tests at the Transportation Security Laboratory (TSL).
    d. Knowledge of the specifics of the ATRs developed by the vendors.
    e. Shield alarm specifications; a shield alarm is a DHS term indicating that a target may not be detected due to insufficient x-ray flux reaching the x-ray detectors.

[18] http://en.wikipedia.org/wiki/Scientific_method

2. More data should have been collected (albeit with expense) to allow testing of ATRs on datasets other than the datasets used for training. However, this would lead to issues related to using the test dataset as follows:
    a. What feedback would be provided to the researchers?
    b. How often could the ATRs be tested?
    c. How difficult would it be to emulate the researchers' computing environments or how to allow researchers access to the test dataset?
3. The inclusion of pseudo targets made the program more difficult to explain to researchers. The scans of the pseudo-targets should have been removed from the dataset.
4. Ground truth labels for non-targets may have reduced the time for developing ATRs.
5. The dataset used in this project may have been easier than the dataset used by vendors because the image quality was better and threats were not scanned with sufficient clutter.
6. There was not enough emphasis on developing features used in the classifier stage of the ATR.
7. The researchers may not have experienced the same pressure that the ATR developers in industry face because people working in the latter group are working for for-profit organizations. That is, better PD/PFA values may be possible with the dataset used for this project.
8. The PD/PFA requirements may have been too easy.
9. The sample ATR (SATR) became the baseline from which the researchers had to perform better. The researchers may have been driven to better detection performance had they started with a sample ATR with better PD/PFA performance.

8 Acknowledgements

The ALERT program management team and the SMEs thank the following people and organizations for their involvement in the segmentation initiative.

- DHS S&T Explosives Division (EXD) for funding ALERT and LLNL to implement this project.
- DHS S&T Office of University Program (OUP) for providing the core funding for ALERT, which supports the ADSA workshops that led to this project.
- Doug Bauer, DHS S&T (retired), Laura Parker, DHS S&T, and George Zarur, DHS (retired), for their vision to involve third parties in the development of technologies for security applications.
- Greg Struba, DHS S&T SETA, for coordinating the participation of DHS at the onset of the program and for reviewing materials during the course of the project.

  David Lieblich, Analogic, David Perticone, L-3 Communications, James Connelly, IDSS, Carl Bosch, Surescan, Piero Landolfi, Morpho Detection, and Christopher Gregory, Smith Detection, for participating in a panel discussion at the program review.

The ATR initiative would not have been a success without the research groups. The content of this report is due mostly to their contributions. We extend our heartfelt thanks to them for their participation and working with the project team to fix issues when they became evident.

9 Project Team

Principal Investigators:
- John Beaty, Northeastern University
- Michael Silevitch, Northeastern University

Program Management:
- John Beaty, Northeastern University
- Ralf Birken, Northeastern University
- Carl Crawford, Csuptwo

Domain experts:
- Carl Crawford, Csuptwo
- David Castañón, Boston University
- Clem Karl, Boston University
- Harry Martz, Lawrence Livermore National Laboratory

Technical editors of final report:
- Carl Crawford, Csuptwo

Preparation of public domain and DHS versions of final report:
- Teri Incampo, Northeastern University

Review of final report:
- Ralf Birken, Northeastern University
- Jens Gregor, University of Tennessee
- Clem Karl, Boston University
- Harry Martz, Lawrence Livermore National Laboratory
- Greg Struba, Booz Allen Hamilton
- Suriyun Whitehead, Booz Allen Hamilton
- Jun Zhang, University of Wisconsin, Milwaukee

Program review logistics and minutes:
- Sara Baier, Northeastern University
- Deanna Beirne, Northeastern University
- Kristin Hicks, Northeastern University
- Teri Incampo, Northeastern University
- Anne Magrath, Northeastern University
- Melanie Smith, Northeastern University

Additional program review minutes:
- Steve Azevedo, Lawrence Livermore National Laboratory
- Harry Martz, Lawrence Livermore National Laboratory

SSI review:
- Horst Wittmann, Northeastern University

SFTP site administrator:
- Clem Karl, Boston University

10 Definitions

10.1 Acronyms

2D	Two-dimensional
3D	Three-dimensional
ADSA	Algorithm Development for Security Applications, a series of workshops conducted at NEU
ADSA01	First ADSA workshop on the check-point application
ADSA02	Second ADSA workshop on the grand challenge for CT segmentation
ADSA03	Third ADSA workshop on whole body imaging (AIT, advanced imaging technology)
ADSA04	Fourth ADSA workshop on advanced reconstruction algorithms for CT-based EDS
ADSA05	Fifth ADSA workshop on fusing orthogonal technologies
ADSA06	Sixth ADSA workshop on specific applications of fusing orthogonal technologies
ADSA07	Seventh ADSA workshop on accelerating development and deployment of advanced reconstruction algorithms for CT-based EDS
ADSA08	Eighth ADSA workshop entitled Automated Threat Recognition (ATR) Algorithms for Explosion Detection Systems
ADSA09	Ninth ADSA workshop entitled New Methods for Explosive Detection for Aviation Security
ADSA10	Tenth ADSA workshop held in May 2014 on air cargo inspection
ADSA11	Eleventh ADSA workshop held in November 2014 on air cargo inspection
AIT	Advanced imaging technology. Technology for finding objects of interest on passengers. WBI is a deprecated synonym.
ALERT	Awareness and Localization of Objects-Related Threats, A Department of Homeland Security Center of Excellence at NEU
AT2	Advanced Technology Two. A DHS designation for scanners that are improved relative to single view line scanners.
ATD	Automated threat detection; a synonym for ATR
ATR	Automated threat recognition or automated target recognition
CCL	Connected components labeling
CDA	Confidentiality disclosure agreement. A synonym of a NDA.
COE	Center of excellence, a DHS designation
CSV	Comma separated values
CT	Computerized tomography
DAS	Data acquisition system
DB	Database (usually an Excel spreadsheet)
dder	Executable name for detection determination scoring program
DE	Domain expert. SME is the preferred term.
DE	Dual energy

DECT	Dual-energy CT. Multi-energy CT (MECT) is a synonym.
DHS	Department of Homeland Security
EDS	Explosives detection system. An EDS is composed of a CT scanner, an ATR algorithm, and a baggage viewing workstation. EDS is also used by TSA to describe certified equipment for detecting threats in checked baggage.
EM	Expectation and maximization algorithm
EXD	Explosive Division of DHS Science and Technology Directorate
FA	False alarm
FBP	Filtered back-projection
FITS	Flexible Image Transport System – a 3D image format used by the software tools
FOV	Field of view
G3D	A program for generating projection data and image for a set of geometric shapes contained in a shape file.
GPU	Graphical processing unit
GT	Ground truth labels
HME	Homemade explosive
HU	Hounsfield Unit (air=-1024, water=0)
ID	Identifier– a physical label with an assigned number affixed to objects scanned in the CT scanner
IQ	Image quality
IRT	Iterative reconstruction technique
LAC	Linear attenuation coefficient
LLNL	Lawrence Livermore National Laboratory
LOD	Level of difficulty
MAR	Metal artifact correction
MBIR	Model based iterative reconstruction
MECT	Multi-energy CT. DECT is a synonym.
MHU	Modified Hounsfield Unit – air = 0 MHU and water=1024 MHU.
MI	Multiple-image - an image format used by G3D, mmi, xpic and other programs.
MMW	Millimeter wave. A type of AIT.
MRF	Markov random field
MTF	Modulation transfer function
MVL	MeVisLab
NDA	Non-disclosure agreement
NEU	Northeastern University
OSR	On-screen resolution
OUP	Office of University Programs – department at DHS S&T
p	Precision
PD	Probability of detection
pdpfa	Executable name for PD/PFA scoring program

PFA	Probability of false alarm
PI	Principal investigator; a synonym of researcher
r	Recall
REAP	Research Evaluation and Advisory Panel. A review process ALERT uses to assure that sensitive material is not contained in publications.
ROC	Receiver operator characteristic
SATR	Sample ATR
satr	Executable name for sample ATR algorithm
SD	Standard deviation
SE	Single energy
SFTP	Secure file transfer protocol (SFTP)
SIRT	Simultaneous image reconstruction technique
SME	Subject matter expert. DE is a deprecated synonym.
SNR	Signal to noise ratio
SOW	Statement of work
SSI	Sensitive security information
SSN	Scan serial number
SSP	Slice sensitivity profile
SVM	Support vector machine
TBD	To be determined
TO1	Task Order 1. The Segmentation Initiative
TO3	Task Order 3. The Reconstruction Initiative
TO4	Task Order 4. This project: the ATR Initiative
TQ	Threat quantity. Minimum mass required by the TSA for detection.
TSA	Transportation Security Administration
TSL	Transportation Security Laboratory
TV	Total variation
XBS	X-ray backscatter. A type of AIT.
XRD	X-ray diffraction
Z	Atomic number
Zeff	Effective atomic number

10.2 Terms

Alarm	A label created (declared) by an ATR. The alarm may be a detection or a false alarm; this determination is performed according to specifications of recall and precision using a ground truth label image.
Algorithm	The mathematical steps (or recipe) used to perform a defined problem. This definition does not include computer code. However, when stated below that an algorithm is a deliverable, the deliverable includes a written description of the algorithms along with source code.
Artifacts	Defects in images such as blurring, streaks, cupping, dishing and noise.
ATR label images	Label images generated by an ATR indicating the presence of alarms.
ATR log file	The log file created by an ATR.
Bag	A bag (or bin) that contains targets and non-targets when scanning on the CT scanner.
Bounding box	The six-tuple of (xmin, ymin, zmin, xmax, ymax, zmax) in image space. The first pixel in a 3D image file is corresponds to xmin=ymin=zmin=1).
Bulk Object	An object that is not a sheet object. Bulk objects include objects of type saline, rubber and clay.
Calculated precision/recall	The precision/recall values calculated for a given label in an ATR label image and a given target in a GT label image. If the calculated recall/precision values are greater than or equal to the target recall/precision values, then the ATR label image is considered a detection.
Certification	A test run at the TSL on checked baggage inspection systems.
Cloud	A multidimensional plot of features for scans of an object in different containers and in different amounts of clutter.
Corrected data	Raw data (projections) after being corrected for scanner and object imperfections, and the logarithm taken.
Correction	A synonym for pre-processing.
Decomposition	Process of using multiple projection sets or images obtained with different x-ray spectra to develop multiple variables for ATR.
Detection	A *detection* occurs when a label declared by an ATR matches the ground-truth for a target, where *match* is defined in terms of recall and precision.
Detection determination log file	The log file created by the detection determination program containing the score information for a single scan, including the number of detections, misses, and false alarms an ATR produced for the given scans.
Digital value	The value of a pixel in an image as stored on a disk. This is not the value of the pixel that is reported by a display program such as ImageJ.
Dimensions	The length of the bounding box in x, y and z. The x and y dimensions are the length of the bounding box in x and y in pixels times the in-plane pixel size. The z dimension is the length of the bounding box in z in pixels times the slice spacing.
False alarm	A false alarm occurs when an ATR produces a label that does not match the requirements for a detection. See Section 2.5 for additional information.
Feature	A characteristic of an object used by an ATR. Examples of features include mass, density and standard deviation.
Ground-truth	See Ground-truth label images.

Ground-truth label images	Label images showing the locations of targets and pseudo-targets in CT scans. These are created with a ground truth generation program.
Height Database	A database (Excel spreadsheet) containing the pixel height of the patient table for each scan.
Image	A 3D set of pixels. The set consists of a set of contiguous 2D slices. The types of image are CT images and label images.
Incomplete Detection	A detection that occurs with relaxed values of precision and recall.
Incumbent vendor	A company developing EDS equipment. The equipment may or may not be deployed in the airports in the United States. The list of incumbent vendors includes L-3 Communications, Reveal Detection, Morpho Detection, Analogic, Rapiscan, Smiths and SureScan.
Inversion	The reconstruction step converting corrected data to reconstructed images. Inversion may be FBP or iterative reconstruction.
Label	A set of non-zero pixels in an ATR label image indicating the presence of an alarm at the corresponding location in the physical bag. A set of non-zero pixels in a ground truth label image indicating the presence of a target at the corresponding location in a bag.
Label Image	An image showing to which label a pixel belongs. Label images are generated by ATRs (ATR label images) as well as a program used to generate ground truth (ground truth label images). An ATR label image indicates the presence of alarms, while a ground truth label image indicates the presence of targets.
Log file	A human-readable output of a program. The suffix of a log file is ".txt" so that it can be opened under Windows.
Meta-data	Data required to reconstruct projections from a scanner. Examples of meta-data include geometric information, readout sequences and file formats.
Miss	A miss occurs when an ATR produces no label for a target in the ground truth label image, or produces a label that does not satisfy the recall/precision specifications for a target in the ground truth label image.
Non-target	An object that an ATR should not detect. If the ATR detects this object, then a false alarm occurs.
Overall detection summary log file	A log file created by the PD/PFA determination program containing info for a given set of scans including the info for each target (taken from the packing database) as well as whether the target was detected or missed. This log file can be opened in a spreadsheet program, such as Excel, so that the info can be sorted and filtered.
Packing database	A database containing information about each object scanned in each bag.
PD/PFA determination log file	A log file created by the PD/PFA determination program containing the overall probability of detection and probability of false alarm for a given set of scans that have already been scored by the detection determination program.
Phantoms	A numerical (mathematical) description of the contents of a bag. Or, a physical piece of luggage containing known geometric shapes.
Post-processing	Image processing that takes place on reconstructed images after the inversion step of reconstruction.
Precision	The fraction of a label declared by an ATR that overlaps with a target as declared by in the ground truth label image. See Section 2.4 for additional information.
Pre-processing	The reconstruction step converting raw data to corrected data.

Pre-processing	The reconstruction step converting raw data to corrected data.
Probability of Detection	For a set of scans, the number of detections divided by number of targets.
Probability of False Alarm	For a set of scans, the number of false alarms divided by the number of non-targets.
Projection data	Collections of line-integrals of objects.
Pseudo-target	A target material with sub-minimum mass or another material with density less than water. A pseudo-target is also a non-target.
Raw data	Projection data directly from the x-ray sensor or data acquisition (DAS).
Recall	The fraction of a target, as declared in the ground truth label image that overlaps with a label detected by an ATR.
Reconstruction	Generation of images from raw data. Reconstruction includes the steps of pre-processing, inversion and post-processing. The resulting images are denoted reconstructed images.
Researcher	A performer for tasks described in TO4. Term is also used in the singular to denote a group of researchers.
Scan serial number	A number from one to the number of bags scanned. This number is used as a unique identifier for each bag scanned.
Security vendor	As synonym of an incumbent vendor.
Segmentation	A step that may be present in an ATR to find the pixels that comprise an object in a scan of a bag.
Shape file	A description of geometric shapes used by G3D.
Sheet Object	A thin object.
Sinogram data	A synonym for projection data.
Slice	A 2D set of pixels corresponding to a 2D cross-section. The set consists of a set of contiguous slices. The types of slices are CT slices and label slices.
Specified precision/recall	The lower bound precision/recall values used in determining whether a label in at ATR label image matches a target in a GT label image.
Spectral CT	CT scanning with a detector having three or more energy bins.
Target	An object that an ATR has to detect. The types of targets are saline, clay and rubber.
Task	A project that is performed as part of this task order. This is a synonym for sub-task.
Task order	A type of funding vehicle that DHS uses for funding performers.
Third-party	A person or group not working for a security vendor. A third-party works in academia or in industry other than the security vendors.
Unfunded participant	A researcher who did not receiving funding from ALERT but was allowed use the database and present at the program review.
xpic	An image display program written by Carl Crawford.
xrec	The Imatron off-line reconstruction program.

11 Supplemental Material in the Appendices

The supplemental material listed in the following subsections is available in appendices to this report. The pathname and filename for each document is provided to access the zip-file provided with the DHS version of this report.

Note that all of the images shown in the supplemental material were obtained from scans on a commercial medical scanner. Explosives and explosive simulants were not scanned. Scans were not obtained on security scanners.

11.1 Specifications

11.1.1 Top Level Spec

"ALERT ATR Project: Top-Level Technical Specifications"

ALERT ATR Project: Top-Level Technical Specifications

Version 5

ALERT ATR Project: Top-Level Technical Specifications, Page 2

Contents

ALERT ATR Project: Top-Level Technical Specifications, Page 3

ALERT ATR Project: Top-Level Technical Specifications, Page 4

1 Preamble

1.1 Executive Summary

The ALERT Center of Excellence at Northeastern University has received funding from the Department of Homeland Security for a project entitled Advances in Automatic Target Recognition (ATR) for CT-Based Object Detection System. This project is also known as Task Order Four and the ATR Initiative. The project addresses improving CT-based explosive detection equipment by developing improved ATR algorithms. The purpose of this document is to provide the technical requirements for the project.

1.2 Scope

This document provides technical requirements for the following aspects of the project.

1. ATR development
2. Portion of the data collection requirements
3. Targets

Technical requirements and details for the software support tools, ground truth labeling techniques, and simulated test images can be found in the additional documentation referenced in Sections 4-6.

1.3 Terms

Term	Definition
Alarm	A label created (declared) by an ATR. The alarm may be a detection or a false alarm; this determination is performed according to specifications of recall and precision using a ground truth label image.
Algorithm	The mathematical steps (or recipe) used to perform a defined problem. This definition does not include computer code.
ATR label images	Label images generated by an ATR indicating the presence of alarms.
ATR log file	The log file created by an ATR
Bag	A bag (or bin) that contains targets and non-targets when scanning on the CT scanner.
Bounding box	The six-tuple of (xmin, ymin, zmin, xmax, ymax, zmax) in image space. The first pixel in a 3D image file is corresponds to xmin=ymin=zmin=1).
Bulk Object	An object that is not a sheet object. Bulk objects include objects of type saline, rubber and clay.
Calculated precision/recall	The precision/recall values calculated for a given label in an ATR label image and a given target in a GT label image. If the calculated recall/precision values are greater than or equal to the target recall/precision values, then the ATR label image is considered a detection
Corrected data	Raw data (projections) after being corrected for scanner and object imperfections, and the logarithm taken.
Correction	A synonym for pre-processing.
Detection	A *detection* occurs when a label declared by an ATR matches the ground-truth for a target, where *match* is defined in terms of recall and precision. See Section 2.4 for more details.

ALERT ATR Project: Top-Level Technical Specifications, Page 5

Term	Definition
Detection determination log file	The log file created by the detection determination program containing the score information for a single scan, including the number of detections, misses, and false alarms an ATR produced for the given scans
Digital value	The value of a pixel in an image as stored on a disk. This is not the value of the pixel that is reported by a display program such as ImageJ.
Dimensions	The length of the bounding box in x, y and z. The x and y dimensions are the length of the bounding box in x and y in pixels times the in-plane pixel size. The z dimension is the length of the bounding box in z in pixels times the slice spacing.
False alarm	A false alarm occurs when an ATR produces a label that does not match the requirements for a detection. See Section 2.5 for additional information.
Ground-truth	See Ground-truth label images.
Ground-truth label images	Label images showing the locations of targets and pseudo-targets in CT scans. These are created with a ground truth generation program.
Height Database	A database (Excel spreadsheet) containing the pixel height of the patient table for each scan. See Section 2.11.3 for additional information.
Image	A 3D set of pixels. The set consists of a set of contiguous 2D slices. The types of image are CT images and label images.
Incomplete Detection	A detection that occurs with relaxed values of precision and recall. See Section 2.6 for additional information.
Label	A set of non-zero pixels in an ATR label image indicating the presence of an alarm at the corresponding location in the physical bag. A set of non-zero pixels in a ground truth label image indicating the presence of a target at the corresponding location in a bag.
Label Image	An image showing to which label a pixel belongs. Label images are generated by ATRs (ATR label images) as well as a program used to generate ground truth (ground truth label images). An ATR label image indicates the presence of alarms, while a ground truth label image indicates the presence of targets.
Log file	A human-readable output of a program. The suffix of a log file is ".txt" so that it can be opened under Windows.
Miss	A miss occurs when an ATR produces no label for a target in the ground truth label image, or produces a label that does not satisfy the recall/precision specifications for a target in the ground truth label image. See Section 2.7 for additional information.
Non-target	An object that an ATR should not detect. If the ATR detects this object, then a false alarm occurs.
Object	A physical item contained in a bag, which may be a target, pseudo-target, or a non-target
Object database	A database (Excel spreadsheet) containing information about each object. See Section 2.11.1 for additional information.
Object form	Objects may be of the following two forms: sheet or bulk
Object type	Objects may be of the following three types: saline, clay, or rubber
Overall detection summary log file	A log file created by the PD/PFA determination program containing info for a given set of scans including the info for each target (taken from the packing database) as well as whether the target was detected or missed. This log file can be opened in a spreadsheet program, such as Excel, so that the info can be sorted and filtered.

ALERT ATR Project: Top-Level Technical Specifications, Page 6

Term	Definition
Packing database	A database containing information about each object scanned in each bag. See Section 2.11.2 for additional information.
PD/PFA determination log file	A log file created by the PD/PFA determination program containing the overall probability of detection and probability of false alarm for a given set of scans that have already been scored by the detection determination program
Phantoms	A numerical (mathematical) description of the contents of a bag. Or, a physical piece of luggage containing known geometric shapes.
Precision	The fraction of a label declared by an ATR that overlaps with a target as declared by in the ground truth label image. See Section 2.4 for additional information.
Pre-processing	The reconstruction step converting raw data to corrected data.
Probability of Detection	For a set of scans, the number of detections divided by number of targets. See Section 2.8 for additional information.
Probability of False Alarm	For a set of scans, the number of false alarms divided by the number of non-targets. See Section 2.9 for additional information.
Projection data	Collections of line-integrals of objects.
Pseudo-target	A target material with sub-minimum mass or a another material with density less than water. A pseudo-target is also a non-target.
Raw data	Projection data directly from the x-ray sensor or data acquisition (DAS)
Recall	The fraction of a target, as declared in the ground truth label image, that overlaps with a label detected by an ATR. See Section 2.4 for additional information.
Researcher	A performer for tasks described in TO4. A synonym of PI.
Scan serial number	A number from one to the number of bags scanned. This number is used as a unique identifier for each bag scanned.
Security vendor	A company developing EDS equipment. The equipment may or may not be deployed in the airports in the United States. The list of security vendors includes L-3 Communications, Reveal Detection, Morpho Detection, Analogic, Rapiscan, Smiths and SureScan.
Shape file	A description of geometric shapes used by G3D.
Sheet Object	A thin object.
Slice	A 2D set of pixels corresponding to a 2D cross-section. The set consists of a set of contiguous slices. The types of slices are CT slices and label slices.
Specified precision/recall	The lower bound precision/recall values used in determining whether a label in at ATR label image matches a target in a GT label image
Target	An object that an ATR has to detect. The types of targets are saline, clay and rubber.
Target form	See “Object form”
Target type	See “Object type”
Task order	A type of funding vehicle that DHS uses for funding performers.
Third-party	A person or group not working for a security vendor. A third-party works in academia or in industry other than the security vendors
xpic	An image display program written by Carl Crawford.

1.4 Acronyms

Term	Definition
2D	Two-dimensional

ALERT ATR Project: Top-Level Technical Specifications, Page 7

Term	Definition
3D	Three-dimensional
ADSA	Algorithm Development for Security Applications, a series of workshops conducted at NEU
ALERT	Awareness and Localization of Objects-Related Threats, A Department of Homeland Security Center of Excellence at NEU
ATR	Automated target recognition
CCL	Connected components labeling
COE	Center of excellence, a DHS designation
CSV	Comma separated values
CT	Computerized tomography
DAS	Data acquisition system
DB	Database (usually an Excel spreadsheet)
dder	Executable name for detection determination scoring program
DHS	Department of Homeland Security
EDS	Explosives detection system. An EDS is composed of a CT scanner, an ATR algorithm, and a baggage viewing workstation.
FITS	Flexible Image Transport System – a 3D image format used by the software tools
FOV	Field of view
G3D	A program for generating projection data and image for a set of geometric shapes contained in a shape file.
GT	Ground truth
HME	Homemade explosive
ID	Identifier– a physical label with an assigned number affixed to objects scanned in the CT scanner
LLNL	Lawrence Livermore National Laboratory
MHU	Modified Hounsfield Unit – air = 0 MHU and water=1024 MHU.
MI	Multiple-image - an image format used by G3D, mmi, xpic and other programs.
MVL	MeVisLab
NEU	Northeastern University
PD	Probability of detection
pdpfa	Executable name for PD/PFA scoring program
PFA	Probability of false alarms
PI	Principal investigator; a synonym of researcher
ROC	Receiver operator characteristic
satr	Executable name for sample ATR algorithm
SME	Subject matter expert
SOW	Statement of work
SSI	Sensitive security information
SSN	Scan serial number
TBD	To be determined
TO1	Task Order 1. The Segmentation Initiative
TO3	Task Order 3. The Reconstruction Initiative
TO4	Task Order 4. This project: the ATR Initiative
TSA	Transportation Security Administration

1.5 Assumptions/Notes

1. This spec will be evolved as necessary.
2. The initial draft is terse by design in order to seed the evolution process.
3. Some tools will not be used by the researchers.

1.6 Document Relationship

This document supersedes the following documents.

[1] Crawford, C. R., "Statement of Work - Advances in Automatic Target Recognition (ATR) for CT-Based Object Detection System – Task Order 4 & ATR Initiative, Version 4, June 12, 2013.

[2] Crawford, C. R., "TO4 (ATR Initiative) Scan Plan," Version 6, October 13, 2013.

[3] Rupcich, F., and Crawford, C. R., "ALERT ATR Project: Software Tools Specifications," Version 5, April 12, 2014.

[4] Rupcich, F. "ALERT ATR Project: Ground Truth Labeling," Version 2, April 12, 2014.

[5] Rupcich, F. "ALERT ATR Project: Simulated Test Images Specification," Version 2, April 12, 2014.

[6] Crawford, C. R., "ATR Project Level of Difficulty Specification," February 5, 2014.

[7] Karimi, Seemeen. "Sample Segmentation Software for Segmentation Grand Challenge," April 30, 2010.

1.7 Background

The Department of Homeland Security (DHS) has requirements for future explosives detection systems (EDS) that include increased probability of detection and decreased probability of false alarm for a larger set of objects and with reduced minimum masses. The larger set of objects includes certain types of homemade explosives (HME). There are indications that these requirements for future EDS equipment may be difficult to achieve with the technologies presently deployed in the field. In order to resolve these issues, DHS has adopted the strategy of augmenting the capabilities and capacities of the vendors of EDS equipment with the involvement of third parties. Third parties are defined as researchers from academia and industry other than the vendors.

DHS has funded ALERT to execute a project denoted the *Automated Target Recognition (ATR) Initiative*, which is also known as *Task Order Four* (TO4). The goal of this project is to involve third parties in the development of ATR algorithms that could eventually be deployed by the incumbent vendors. The work will be led by the Northeastern University component of the ALERT DHS Center of Excellence (COE). The investigators for the projects will be comprised of researchers both within and outside of the current group of people being supported by the COE.

The research is designed with the following outcomes for DHS.

- The program will improve ATRs. The improved target recognition may lead to decreased minimum target mass, increased target population coverage, increased probability of detection and decreased probability of false alarm.
- The program will increase involvement of third parties via the availability of common CT datasets, and tools, which will increase the work in target recognition, and the number of students who can join the workforce of the vendors and DHS.
- The program will foster collaboration between academics, national laboratory personnel and incumbent security industry vendors.

Technical interchange will be facilitated near the end of the project so that the researchers can present their results to the security vendors, DHS and other third-parties. The results will be documented in a final report for DHS.

1.8 Overview of Project

An overview of the project is described in this section.

1. Bags will be packed to represent what is found in stream of commerce in airports.
2. The target materials will be saline, polymer (modeling) clay, and rubber.
3. Targets of any material may be considered “bulk” or “sheet” form. The rule-set for determining whether a target is a bulk or a sheet is described in Section 2.10.3.
    a. Most rubber targets are sheets (exception: rubber rods)
    b. Most saline and clay targets are bulk
4. The bags will be scanned on a medical CT scanner resulting in images of the bags. The scans will be single energy.
5. The contents and their placement in bags will be documented.
6. The voxels corresponding to the targets in the images will be marked and stored in label images. This marking is known as ground truth.
7. ATRs will be developed using the images and the ground truth.
8. The ATRs will be assessed using the following criteria.
    a. Minimizing the probability of false alarm (PFA) for a specified probability of detection (PD)
    b. Minimal use of algorithms for specific target configurations (known as corner cases)
    c. Minimal overtraining on test data
    d. Novelty compared to the prior art
    e. Ability to detect targets in difficult configurations
    f. Potential to be extended to detect additional targets
9. Software will be supplied to compare the results of ATRs with the ground-truth.
10. The following items will be supplied to assist the development of ATRs
    a. A sample (notional) ATR so that common functions (e.g., reading and writing images and results) do not have to be replicated by each PI. The benchmark ATR may be updated as necessary during the course of the project.

ALERT ATR Project: Top-Level Technical Specifications, Page 10

b. The segmentation algorithms developed for the Segmentation Initiative (Task Order 1) may be available for use in this project. Some of the algorithms may be patented.
c. A bibliography describing related prior art in the ATR field.

2 General Requirements

2.1 CT Scanning

See the TO4 (ATR Initiative) Scan Plan for more details pertaining to the scanning specifications and process.

2.1.1 Scan Characteristics

The CT scans will have the following characteristics.

1. CT scanner: Imatron C-300
2. Image size: 512 x 512
3. Field of view: 475 mm
4. In plane pixel size: (475/512) = 0.928 mm
5. Slice spacing: 1.5 mm
6. Pixel volume: (475/512)^2 * 1.5 = 1.291 mm^3
7. Digital values: air = 0, water = 1024
8. Minimum pixel value: 0 MHU
9. Maximum pixel value: 32,767 MHU
10. File format: FITS (16-bit, unsigned integer)

2.1.2 CT Scanner Axes

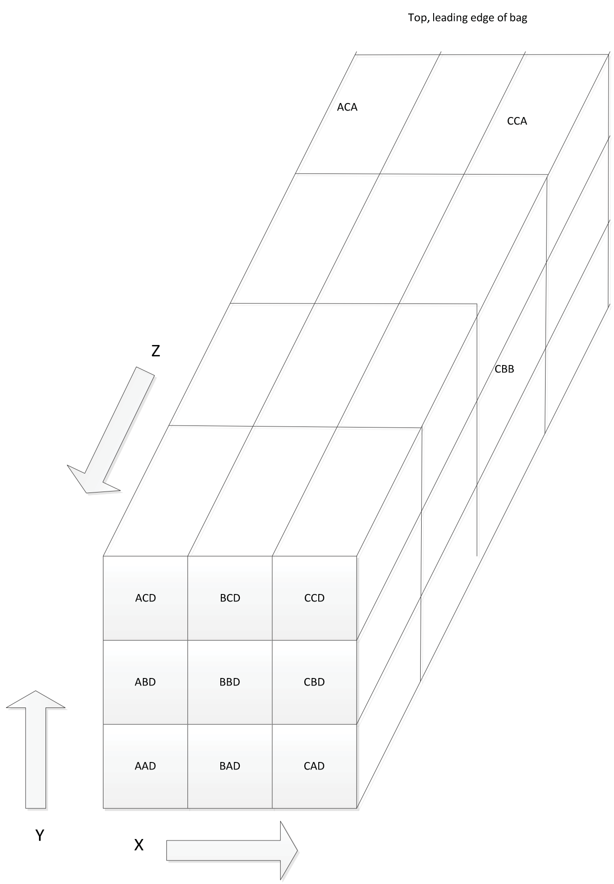
The following axes shall be used for the CT-scanner

1. x: horizontal axis of axial slice
2. y: vertical axis of axial slice
3. z: parallel to direction of table movement for helical scans

2.1.3 Location Code for Objects Placed in a Bag

A three-letter code is used to note where objects are placed in a bag. The code is of the form xyz, where x, y, and z are letters showing the location along the x, y and z, axes, respectively. The x- and y-axes are split into three sections denoted A, B, and C. The Z-axis is split into four sections denoted A, B, C and D. The following diagram shows some are the location codes map to a bag.

ALERT ATR Project: Top-Level Technical Specifications, Page 11



2.1.4 Preferred Axes for Objects

1. Cylinders: axis of rotation
2. Sheets: Parallel to conveyor belt
3. Cuboids: Longest dimension

2.1.5 Orientation Codes for Objects

The orientation code is used to specify how the preferred axis of an object is oriented in bag. The values of the code are as follows.

1. Aligned to an axis:
   a. X: aligned to x axis
   b. Y: aligned to y axis
   c. Z: aligned to z axis
2. Not aligned to an axis but in a plane aligned with two axes:

ALERT ATR Project: Top-Level Technical Specifications, Page 12

   a. XY: in xy-plane
   b. YZ: in yz-plane
   c. XZ: in xz-plane
3. Other
   a. N: not aligned with an axis and not in plane aligned with two axes

Notes
1. A plus sign (+) sign or minus sign (-) shall be appended to all the orientation codes to show how the preferred axis of an object.

2.1.6 IDs

1. IDs for targets have to be unique and numeric
2. Each packing or shape of an object has own ID. For example,
   a. Each bottle of saline has its own ID
   b. Each shape (cutting) of a rubber sheet has its own ID
3. The bulk (source) material(s) for targets should also have unique IDs. For example, the box of modeling clay should be given an ID. Each time a piece of clay is cut from the bulk or a piece is molded, it should be given a new ID.

2.2 Images and Files

2.2.1 Label Images

A label image is a 3D image that indicates if a pixel in a CT image corresponds to a target. A label image is output from a program that generates ground-truth or by an ATR.

1. Size: same (# slices = N , # rows = 512, # columns = 512) as CT image from which labels are generated
2. Sources: ATR program, GT Generator program
3. Background label: digital value of 0
4. Foreground values:
   i. ATR label images: positive integers assigned by ATR
   ii. GT label images: ID of targets
   iii. Maximum value: 65535 (maximum value of unsigned short int)
   iv. Labels within a label image do not have to have sequential values (e.g., 1, 2, … N, where N is the number of labels).
5. General: A label image pixel can be assigned to only one label
6. Maximum number of labels: 100

2.2.2 File Compression

1. Files: CT images or label images
2. Algorithms: gzip
3. Application: mandatory

ALERT ATR Project: Top-Level Technical Specifications, Page 13

2.2.3 Image Formats

2.2.3.1 Images

1. Format: 3D FITS, 16-bit, unsigned integer
2. Types: CT images and label images
3. File suffix: .fits

2.2.3.2 Log Files

1. Format: Windows text format; compatible with notepad
2. Types:
    a. ATR log file
    b. Detection determination log files
    c. PD/PFA log files
3. File suffix: .txt, .xls

2.2.4 Databases

1. Format: CSV (derived from Excel spreadsheets)
2. Types: packing database, object database, height database
3. File suffix: .csv (for software tools), .xls (human readable form)

2.2.5 File Naming Conventions

Filenames are a single letter followed by the SSN (zero padded to three digits), followed by the extension *fits.gz* (gzipped compressed FITS format). The letter code is as follows:

I – CT image
G – GT label image
A – ATR label image

For example, the CT, GT label, and ATR label images for SSN 50 are I050.fits.gz, G050.fits.gz, and A050.fits.gz, respectively.

NOTE: The SSNs, and thus the filenames, range from 004 to 193. However, due to corrupt/missing data, SSNs 27 and 160 are not used.

2.3 Project FTP Site

1. URL: eng-filetransfer.bu.edu/eng_research_TO4
2. Directory structure: TBD

2.4 Detection

A detection occurs when an alarm declared by an ATR *matches* the ground-truth for a target. The term *match* is defined in terms of *recall*, *R*, and *precision, P*. Let *G* correspond to the set of pixels in the ground-truth for a target. Let *S* correspond to the set of pixels declared to be an alarm by an ATR. Then recall and precision are defined as follows.

$$R = \frac{volume(G \cap S)}{volume(G)}$$

$$P = \frac{volume(G \cap S)}{volume(S)}$$

A detection occurs when:

1. For bulk objects: $R \geq 0.5$ and $P \geq 0.5$.
2. For sheet objects: $R \geq 0.2$ and $P \geq 0.2$

2.5 False Alarm

A false alarm occurs when an ATR creates a label that does not meet the requirements for a detection.

2.6 Incomplete Detection

An incomplete detection is a detection that occurs for the values of precision and recall, as shown in Section 2.4, multiplied by the factor, *alpha*. The default value of alpha is 0.0, which implies that an ATR label that intersects a GT label by at least one pixel meets the requirement for an incomplete detection. However, if an ATR label meets the requirements for a detection for a GT label, then it will not be counted as an incomplete detection for that GT label.

Note the following:

- Incomplete detections ***do not*** count as a detection
- Incomplete detections ***do*** count as false alarms

2.7 Miss

A miss occurs when an ATR produces no label that satisfies the precision and recall specifications for a target in the ground truth label image.

2.8 Probability of detection (PD)

Probability of detection is defined as the number of detections (see Section 2.4) divided by the number of targets present in a set of scans. The set of scans may be less than all the bags in the packing database. There may be different types of PDs for different sets of targets. The following types of PD may be used:

1. For all targets
2. For each type of target
3. For different levels of clutter
4. For different orientations
5. For different locations
6. For combinations of the above

2.9 Probability of false alarm (PFA)

Probability of false alarm is defined as the number of false alarms (see Section 2.5) divided by the number of non-targets for a given set of scans.

ALERT ATR Project: Top-Level Technical Specifications, Page 15

2.10 Targets

2.10.1 Materials

1. Saline doped to have a densities overlapping with other liquids commonly found in bags.
2. Modeling clay (polymer)
3. Rubber

2.10.2 Mass

1. Minimum mass: 250 g
2. Maximum mass: None

2.10.3 Thickness (Sheet vs. Bulk)

The following rule-set is used to determine whether an object is a sheet pseudo-target, a sheet target or a bulk target:

thickness < 1/4 " : sheet pseudo-target
1/4" <= thickness <= 3/8" : sheet target
thickness >3/8" : bulk target

2.10.4 Pseudo-Targets

1. Pseudo-targets (PT) are one of the following types
   a. Target materials listed in Section 2.10.1 with masses ≥ 125 g and < 250 g
   b. Targets materials with a thickness less than ¼"
   c. Powders with masses greater ≥ 125 g and density < 1 g/cc
2. ATRs are not required to detect PTs. A detection on a PT will not be considered to be a false alarm.

2.11 TO4 Database

The TO4 database is a version-controlled Excel workbook comprising multiple worksheets containing information about each object and scan. The overall database is further broken down into three individual databases (each is a single spreadsheet in the TO4 database workbook):

1. Object database – contains information about each object
2. Packing database – contains information about each bag
3. Height database -- indicates the height of the patient table for each bag

The software tools use CSV formatted versions of the three database files. The CSV versions of the database files are distributed with the tools package. More detailed specifications of the databases can be found in the ALERT ATR Project: Software Tools Specifications.

2.11.1 Object Database

The object database contains information about each object (both targets and non-targets) that was scanned for this project, including the ID, object description, material type and form, mass, volume, and dimensions.

ALERT ATR Project: Top-Level Technical Specifications, Page 16

2.11.2 Packing Database

The packing database contains information about each object in each scan, including the ID, location code, orientation code, level of difficulty, and bounding box.

2.11.3 Height Database

The height database contains the pixel height of the patient table for each scan. It is used by the sample ATR algorithm to zero out the pixels below the patient table for a given CT image.

3 ATR Specification

3.1 Owners

1. Jens Gregor, University of Tennessee
2. Synho Do, Massachusetts General Hospital
3. Charles Bouman et al., Purdue University
4. Jun Zhang, University of Wisconsin, Milwaukee
5. Unfunded participants

3.2 Synopsis

1. Detects targets in sets of CT slices.

3.3 Arguments

1. Filename of input FITS CT image
2. Filename of output FITS image label image [default: derived from input filename]
3. Filename of output log file [default: derived from input filename]

3.4 Inputs

1. 3D FITS containing CT images

3.5 Functions

1. Detect targets
2. Determine the following features for each label
    a. Mass
    b. Volume
    c. Density (mean and standard deviation)
    d. Number of voxels
3. Create log file

3.6 Outputs

1. 3D FITS containing label image
2. ATR log file – specified Section 7.

3.7 Deliverables

1. Algorithm description

ALERT ATR Project: Top-Level Technical Specifications, Page 17

2. Log files
3. Label images

3.8 Acceptance Criteria

1. Maximize PD and minimize PFA

3.9 Notes

1. The ATR may be based on the sample ATR, satr.c.
2. The bags are scanned at the same height. Therefore, pixels after a TBD tow may be zeroed.
3. Multiple researchers will be developing ATRs. The researchers shall develop their algorithms independently.
4. The input to the ATR shall be images. Projection data (raw and corrected) may be used by the ATR but not reconstructed.
5. There shall not be an upper limit on mass or volume.
6. Detection should be independent of shape, size, location, orientation, clutter, and concealment. This means that the researcher should not try to meet the PFA requirement by not detecting configurations of targets that lead to high false alarms.
7. PD may be weighted to emphasize targets whose images are corrupted by CT artifacts.
8. The researchers are requested to:
    a. Separate the data into training and test sets
    b. Not over-train on the data
    c. Design their ATRs to be extensible so that additional targets can be considered in the future.
9. Containers (e.g., bottles) for liquids are not considered to be part of the target.
10. The ATRs shall be different than the methods presented in the prior art.
11. There is no requirement to report the type of target (e.g., saline, modeling clay or rubber sheet).

4 Software Tools Specification

The following software support tools were created to standardize both the scoring of the ATR algorithms and the reporting of PD/PFA. Details can be found in ALERT ATR Project: Software Tools Specifications.

1. Sample ATR
2. Detection determination (scoring)
3. PD/PFA determination
4. Generate PD/PFA
5. GT verification
6. MI to FITS file converter
7. FITS to MI file converter
8. Raw to FITS file converter
9. Merge CT and label images

ALERT ATR Project: Top-Level Technical Specifications, Page 18

5 Ground Truth Labeling

Semi-automated segmentation and labeling each target was performed to obtain ground truth label images. Details of the ground truth labeling process can be found in ALERT ATR Project: Ground Truth Labeling.

6 Simulated Test Images Specification

Simulated test images were generated for testing ATRs. Details can be found in ALERT ATR Project: Simulated Test Images Specification.

7 ATR Log File Format

7.1 Synopsis

The purpose of this section is to specify which information should be supplied in the ATR log file when an ATR program processes a set of images. The format of the information is also specified.

7.2 Format

1. One ATR log file per scan.
2. Contains information about processing one set of CT slices.
3. Contains information for all labels including the background label.
4. File should have .txt suffix and be readable by *notepad* in Windows.
5. Information is supplied as *[keyword] value* with an optional *(units)* inserted after the keyword.
6. Whitespace can be added.
7. Keywords are not case sensitive.
8. Pixel indices begin with (1,1,1).
9. The order of keywords has to match table indicated the following section.
10. If a keyword is not applicable, then its value should be left blank.
11. The symbol > means tag repeats.
12. MHU are modified Housfield Units (HU). Air and water are 0 MHU and 1024 MHU, respectively.
13. The volume of a pixel is specified in Section 2.1.1.
14. <space> means insert blank line in an ATR log file at this point.
15. See below for a sample log file.
16. Include filename suffixes when filenames are reported.
17. All text after a pound sign (#) is considered to be a comment or additional information supplied by an ATR.

7.3 Information

Keyword	Units	Contents	Remarks
[Performer]		Name of institution or researchers or both	
[Date]		Date images processed	May include time.
[Time]		Time images processed	May include date.
[Image-name]		Image filename	Set of CT slices

ALERT ATR Project: Top-Level Technical Specifications, Page 19

[Image-format]		Format of input image file used	FITS is the only accepted format.
[CT-columns]		Number of columns per image	
[CT-rows]		Number of rows per image	
[CT-slices]		Number of 2D images in the input file	
[CT-first]		First slice used in the set of 2D slices.	Should be one if all the input slices are used.
[CT-count]		Number of 2D slices processed.	Should be equal to the [CT-slices] if all the slices are used.
[CT-fov]	(mm)	Scanner FOV	
[CT-pixel]	(mm)	Pixel size	
[CT-slice-space]	(mm)	Slice spacing	
[CT-offset]	(MHU)	Value subtracted to make the CT value of air equal to 0 MHU.	
[CT-dimension-z]	(mm)	[CT-count] * [CT-slice-space]	
[CT-mean]	(MHU)	Mean of all pixels sent to the ATR.	
[CT-mass]	(g)	Mass of all the pixels sent to the ATR.	Mass is estimated using: CT/1024 * voxel size
[Label-name]		Filename of output label image	
[Label-format]		Format of output label image	FITS is the only accepted format.
[OS]		Operating system	E.g., Windows or Linux
[Executable]		filename of executable program	
[Version]		Version # of ATR algorithm	
[Total-labels]		Total number of labels detected	Excluding background
<space>			
>[Label-num]		Label number	1. From 0 to total-labels 2. 0 is for the background 3. This and remaining keywords are repeated for each of the labels segmented 4. Calculations performed on CT pixels are limited to the CT pixels that have the indicated label value.
>[Label-id]		Value of label in the label image	Does not have to match target ID

ALERT ATR Project: Top-Level Technical Specifications, Page 20

>[Slice-first]		First slice containing label	This and the next five keywords are for a rectangular bounding box in the image coordinate system
>[Slice-last]		Last slice containing label	
>[Row-first]		First row containing label	
>[Row-last]		Last row containing label	
>[Column-first]		First column containing label	
>[Column-last]		Last column containing label	
>[Dimension-x]	(mm)	(Column-last – Column-first+1) * pixel size	
>[Dimension-y]		(Row-last – Row-first+1) * pixel size	
>[Dimension-z]		(Slice-last – Slice-first +1) * slice spacing	
>[Voxels]		Voxels segmented – that is, number of voxels with value *Label-id*	
>[Mass-CT]	(g)	Mass of label using CT values	Mass is estimated using: CT/1024 * voxel size
>[Volume]	(cc)	Volume of label	
>[Density –CT]	MHU	Sum of values of the voxels divided by the number of voxels detected	
>[Density-std-CT]	MHU	Standard deviation of the values of the CT voxels	
><space>			

7.4 Sample ATR Log File

[Performer] Carl Crawford, Csuptwo
[Date]: Thu Dec 12 13:23:25 2013
[Time]: Thu Dec 12 13:23:25 2013
[CT-name] I076.fits
[CT-format] FITS
[CT-columns] 512
[CT-rows] 512
[CT-slices] 270
[CT-first] 1
[CT-count] 270
[CT-fov] (mm) 475.00
[CT-pixel] (mm) 0.93
[CT-slice-space] (mm) 1.50
[CT-offset] (MHU) 0
[CT-dimension-z] (mm) 405.00

ALERT ATR Project: Top-Level Technical Specifications, Page 21

[CT-mean] (MHU) 22.22
[CT-mass] (g) 1982.50
[Label-name] label.fits
[Label-format] FITS (16-bit unsigned short)
[OS] Linux
[Executable] satr
[Version] $Id: satr.c,v 1.1 2013/10/18 13:55:33 franco Exp franco $
Total-labels includes label (0) for background
[Total-labels] 2

**** satr program variables ****
#min mass (g) = 50.00
#low threshold (MHU) = 1000
#high threshold (MHU) = 2000
#ccl delta (MHU) = 100
#connectivity = 0

Label-num=0 is the background
[Label-num] 0
[Label-id] 0
[Slice-first] 1
[Slice-last] 270
[Row-first] 1
[Row-last] 512
[Column-first] 1
[Column-last] 512
[Dimension-x] (mm) 475.00
[Dimension-y] (mm) 475.00
[Dimension-z] (mm) 405.00
[Voxels] 69751026
[Mass] (g) 459.12
[Volume] (cc) 90051.12
[Mean] (MHU) 5.22
[Standard-deviation] (MHU) 83.69

[Label-num] 1
[Label-id] 1
[Slice-first] 1
[Slice-last] 72
[Row-first] 185
[Row-last] 350
[Column-first] 163

[Column-last] 374
[Dimension-x] (mm) 196.68
[Dimension-y] (mm) 154.00
[Dimension-z] (mm) 108.00
[Voxels] 382970
[Mass] (g) 578.37
[Volume] (cc) 494.43
[Mean] (MHU) 1197.85
[Standard-deviation] (MHU) 120.40

8 Revision History

Version	Date	Author	Revisions
1	12/2/2013	CRC	Initial release of top-level spec. Based on latest versions of scanning spec, SOW and tool-spec.
2	12/12/2013	CRC + FJR	Changes based on in part based on FJR's editing and Jens's comments.
3	4/6/2014	CRC, FJR	Additional revisions. Moved SW Tools specifications to SW Tools Spec document.
4	4/12/2014	CRC, FJR	Reference other documents. Re-organized.
5	5/24/2014	FJR	Further defined "Incomplete Detections"

11.1.2 Level of Difficulty

"ATR Project Level of Difficulty Specification"

ATR Project Level of Difficulty Specification

Synopsys

The purpose of this document is to define the level of difficulty (LOD) that will be applied to each instance of a target scanned for the ATR Project (Task Order 4).

Assumptions

1. The LOD are as follows:
    a. Low (L)
    b. Medium (M)
    c. High (H)
2. Only Low and High will be used for this project at this time
3. Pseudo targets (PT) are considered to be targets when determining LOD
4. The LOD is applied to each target after it is packed and scanned. A specific packing is denoted a "configuration" in this document.
5. The LOD will be noted in the project database (spreadsheet) in the tab named "packing".
6. The LOD may be used to weight PD more heavily for configurations with high LOD.

LOD Specification

A configuration will be assigned a low LOD by default. A high LOD will be assigned if one of the following conditions is satisfied.

1. The target is difficult to segment from its surroundings
2. The features of the target are corrupted
3. Detecting the target will lead to a h PFA

What follows are reasons for these conditions to occur.

1. Target is touching a non-target with similar density
2. A *discontinuous* shape was created because of object philosophy. This subjective term means that the target could have been assembled from two more pieces of the base material (saline, rubber and clay).
3. Streaks pass through the target
4. Low frequency shading (caused by beam hardening and scatter) depress the CT values by more than 10%
5. Sausage-like scanned with its long axis contained in the x-y plane.
6. Sheet targets scanned with most of its mass contained in the x-y plane.
7. Sheets are rolled up
8. Texture is present in the target. For example, clay mixed with glass beads.
9. High electronic or quantum noise present in the CT images
10. All PTs

11. Saline with concentration of less than ~7.5 grams of salt / 100 grams of solution

Notes

1. LOD is applied per configuration, not slice-by-slice
2. The assigned LODs are subject to change
3. LOD will be biased because difficulty may be influenced by previous experience developing ATRs.

11.2 Programmatic

11.2.1 Kickoff Presentation

"ATR Initiative (Task Order 4): Technical Kickoff Meeting"

ATR Initiative (Task Order 4): Technical Kickoff Meeting

August 22, 2013
Revised, October 9, 2013
Revised, October 10, 2013
Revised, November 4, 2013
Revised, December 16, 2013

1

Bottom Line

- Develop automated target recognition (ATR) algorithm to detect targets in scans on a medical CT scanner
- Input = 3D CT data + projections
- Output = 3D label image (i.e., a "mask") indicating pixels of detected targets
- Targets
  - Saline, modeling clay, rubber sheets
- Detection of objects in ATRs
  - Defined in terms of recall and precision
  - Determined using a ground truth label image
- Scoring of ATRs
  - Probability of detection (PD) > x%
  - Probability of false alarm (PFA) < y%
  - TBD function of PD and PFA (F-score?)
  - PD and PFA may be defined per object, not per bag/target

2

Bottom Line (II)

- Prior art methods are proprietary and classified
  - Will not know if results are better
- Success
  - Understanding ATR problem
  - Being able to work with suppliers in the future
  - Solving detecting difficult cases more important than trying to detect easy cases
- Sample ATR, scoring programs will be supplied to reduce development efforts
- Bibliography supplied of prior art

3

Vernacular

- ATR (automated target recognition) = ATD (automated target detection)
- Targets = objects of interest
- Project names
  - ATR project or initiative
  - Task Order 4 (TO4)

4

Team - PIs

- ATR
  - Charlie Bouman – Purdue University
  - Synho Do – Massachusetts General Hospital
  - Jens Gregor – University of Tennessee
  - Jun Zhang – University of Wisconsin, Milwaukee
- Data collection
  - Doug Boyd, Telesecurity Sciences
  - Rick Moore, Alyssa White, Massachusetts General Hospital
- Tools and ground truth
  - Franco Rupcich, Independent consultant
- Technical leadership
  - David Castanon, Clem Karl, Boston University
  - Carl Crawford, Csuptwo
  - Harry Martz, Lawrence Livermore National Laboratory
- Programmatic leadership
  - Michael Silevitch, John Beaty, Northeastern University/ALERT

5

ATR Definition

- Input to ATR: 3D CT images + projections
  - Objects are not segmented
  - Feature are not extracted
- CT images:
  - Use supplied images
  - Do not reconstruction projections to create new images
- Possible Functions: segmentation, feature extraction, CT correction, classification
- Outputs from ATR: red/green light, label images, log file

6

ATR Implementation

- Sample ATR will be supplied based on Karimi code for Segmentation Project
  - Demonstrate reading FITS images, writing label images and log files
  - Can rip out algorithm and insert your own
  - C/Linux
- Execution time out of scope
- Do not over-train on the data or use illegal features
  - Will have to discuss algorithms and training methods

7

ATR Overview (Prior Art)

Multi-Path Segmentation → Feature Extraction → CT Correction → Classifier

Multi-Path Segmentation	Feature Extraction	CT Correction	Classifier
Sheet filter/path Bulk filter/path Weapons filter/path	Mass Density Z-effective Texture Volume	Orientation Resolution	Non-linear thresholds

Do something other than prior art

8

Features (Prior Art)

- Mass
- Mean: LAC, Zeff
- Standard deviation: LAC, Zeff
- Histograms
- Higher-order moments
  - Skew, kurtosis, entropy
- Texture
  - Wavelets

Your responsibility to determine relevant features

9

Illegal Features

- Shape (except for minimum sheet thickness)
- Size
- Orientation
- Location
- Maximum mass, volume
- Container type

Will review your features during course of the project

10

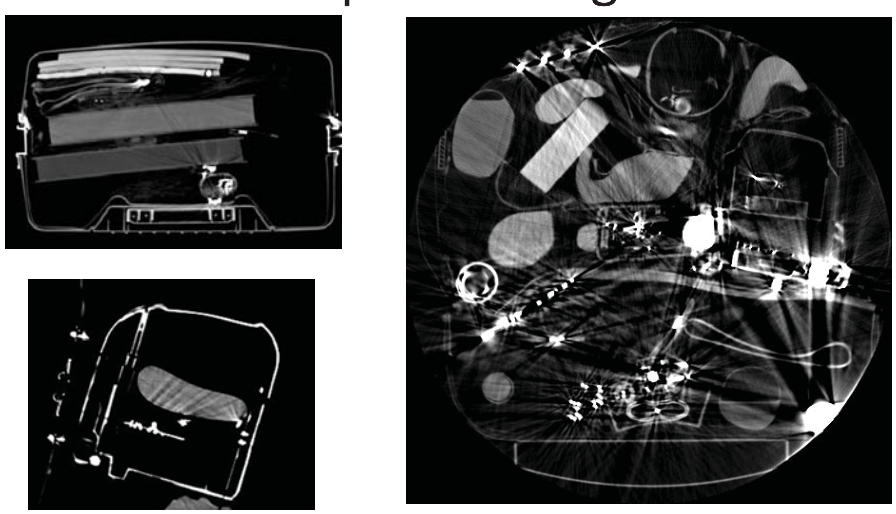
CT Artifacts

- Finite and spatially dependent resolution
- Streaks
- Additive noise
- Rings, bands
- Low-frequency shading (cupping, dishing)
- CT number shifts

Literature says artifacts reduce detection performance

11

Sample CT Images



Label Images

- Label Image: *An image generated either by an ATR or by a ground truth generating program indicating to which **label** a pixel belongs*
- Label:
  - *A set of non-zero pixels in an **ATR label image** indicating the presence of an alarm at the corresponding location in the physical bag/bin*
  - *A set of non-zero pixels in a **ground truth (GT) label image** indicating the presence of a target at the corresponding location in the physical bag/bin*
- The number of pixels in a label image is the same as in its corresponding CT image (*N*x512x512)
- A label image pixel can be assigned to only one object
- Values:
  - 0 = background
  - Label/tag for object for ground truth
  - >0 = object number from ATR
    - Does not have to match object tag

13

Targets

- Saline: ~1035 – 1150 MHU
  - Container not part of target; only the saline
- Modeling (polymer) clay
- Rubber sheers: ¼” thickness (minimum)
- Minimum mass: 250 g (physical, not CT)
- Maximum mass: none
- Targets will not be split; they will be contiguous

14

Pseudo Targets

- Saline, clay, rubber sheets
  - 125 g < mass < 250 g
- Powders:
  - Density < 1 g/cc
  - Mass > 125 g
- Sheets
  - < ¼” thick
- Used to test lower mass and density thresholds
- Will not count in PD or PFA calculations

15

Packing

- Targets packed with different
  - Shapes
  - Containers (not part of target)
  - Concealment
  - Clutter
  - Location
  - Orientation
- An indication of level of difficulty of detection will be supplied
- Targets will be scanned without clutter (bare) to show clean images

16

Scanning & Data Sets

- Scan on Imatron medical CT scanner
  - Same scanner and protocol as reconstruction project
  - Single energy
  - Raw data collected, but not be used on this project
  - Date ~9/30
  - Scan spec distributed soon
    - Send feedback on what needed to scan
- ~200 scans
- Ground truth will be created only for targets and pseudo targets, not for non-targets
  - Manually created to prevent CT artifacts from splitting objects
- Use TO3 data for now

17

Database

- CT scans
- Ground truth
- Information on packing
  - List of objects and characteristics (mass, size, etc.)
  - Pictures
  - Videos
  - Objects have IDs

18

ATR Testing

- Run by performer at their site
  - Too difficult to execute at ALERT
  - Honor system not to over-train
- Results scored by provided tools
  - PD/PFA
  - [optional] ROC area
  - Some indication of difficult cases

19

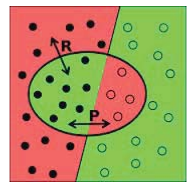
Performance Metrics

- PD > x, PFA < y
  - Values TBD
  - PD based on # targets scanned
  - PFA based on # bags scanned
  - TBD metric combining PD and PFA
- Number of FAs/bag
- [optional] area under ROC
- PD
  - May be weighted for difficult cases
  - Report for each type of target and average

20

Detection and False Alarm

- A detection occurs when an object declared by an ATR *matches* the ground-truth for a target. The term *match* is defined in terms of *recall*, *R*, and *precision, P*. Let *G* correspond to the set of pixels in the ground-truth for a target. Let *S* correspond to the set of pixels declared to be an object by at ATR. Then recall and precision are defined as follows.



- $R = \frac{volume(G \cap S)}{volume(G)}$
- $P = \frac{volume(G \cap S)}{volume(S)}$
- For this project a detection occurs when *R*≥0.5 and *P*≥0.5.
  - May relax for all objects depending on accuracy of ground truth

21

More Performance Criteria

- Minimize use of special cases (corner cases)
- Feature space chopped up
  - Over-training
- Extensible for new targets

22

File Formats

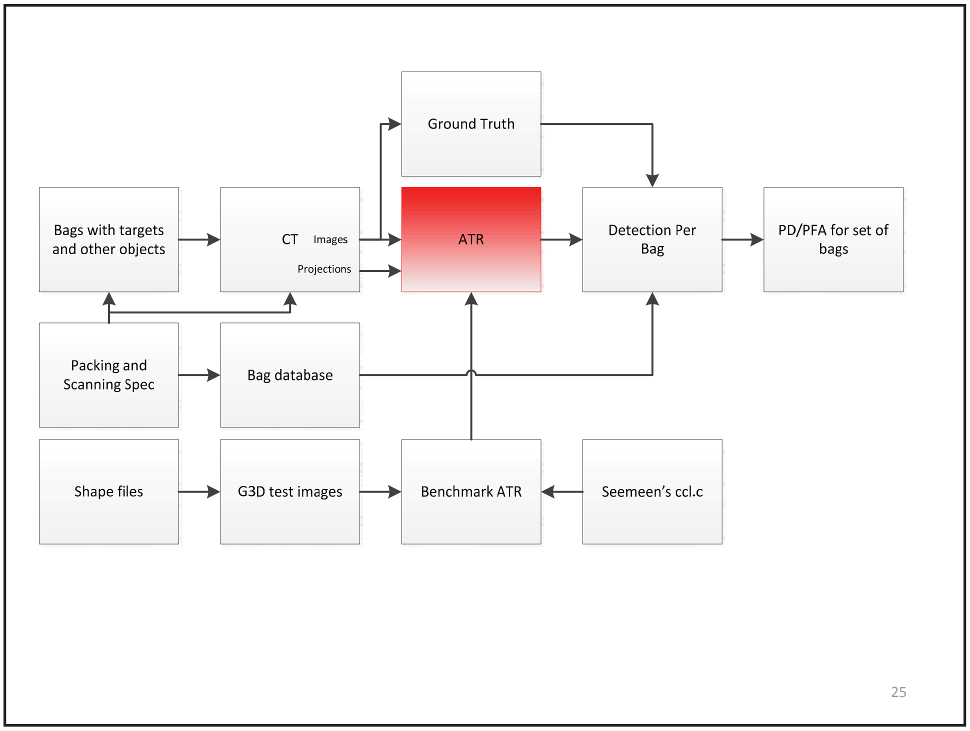
- FITS (Flexible image transport system)
  - 3D
  - CT images converted from DICOM to FITS
  - Code in C/Matlab written by Jens
  - http://fits.gsfc.nasa.gov/
  - Supported by Imagej

23

Support Functions (Tools)

- Franco Rupcich coding in C
- Sample ATR
  - Reading image
  - Writing results (label, log files)
  - Revised as necessary
  - Replace ATR functions with your own
- Scoring software
  - Detection using recall/precision
  - PD/PFA
- Simulated images for validation
- Image conversion to FITS
- Ground truth - Mevislab

24



Reasons for This Process

- Detection requirements are classified
- Data from deployed equipment are SSI or classified, and are under export control
- There is no publicly available set of images that are representative of challenging ATR problems for explosive detection systems.
- The business interests of the vendors should be protected
- DHS/TSA policies do not allow TSL to test components (e.g., an ATR) separate from a complete scanner
- There may be privacy concerns with scans on AIT equipment.

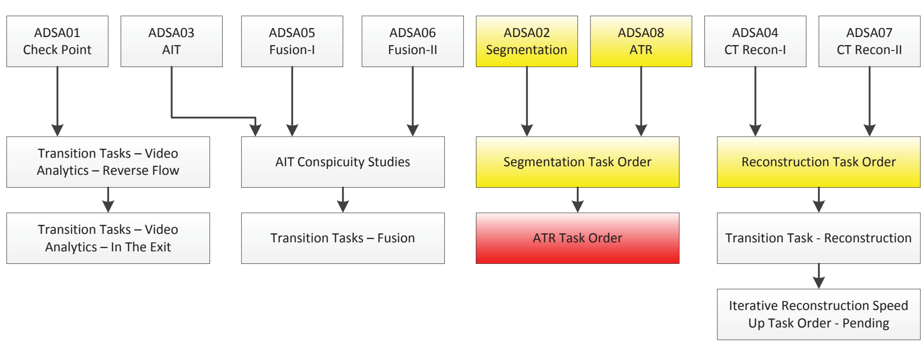
26

Program Success

- The program will improve ATRs. The improved target recognition may lead to decreased minimum target mass, increased target population coverage, increased probability of detection and decreased probability of false alarm.
- The program will increase involvement of third parties via the availability of common CT datasets, and tools, which will increase the work in target recognition, and the number of students who can join the workforce of the vendors and DHS.
- The program will foster collaboration between academics, national laboratory personnel and incumbent security industry vendors.

27

ADSA – Task Order Linkage



28

Uniqueness

- Bibliography and copies prior art provided
  - Patents, presentations, reports, articles
  - Vendors may be doing something else
- Do not replicate prior art
- Extend sample ATR

29

Collaboration

- Researchers encouraged to collaborate with other team members
  - Tools, image formats
  - Not on algorithms
- Segmentation expertise from segmentation project

30

Schedule

- Contracts written for one year to ~8/2014
- ALERT will request no cost extension from DHS to ~2/2015
- Scanning on 9/30
  - Use TO3 data for now
- ~6 weeks before end of project
  - Technical interchange
  - Final reports due
- Request constant progress

31

Status Reporting

- Monthly tele-conferences
  - First Thursday of month, Noon ET
- Monthly status reports
  - Template distributed
  - Acceptable to distribute reports to team?
  - Due first Thursday of month
- Calls/visits from technical leadership
  - Crawford, Castanon, Martz
- Miscellaneous emails

32

Out of Scope

- Execution speed + computational expense
- Equating to performance of equipment deployed by TSA
- Reconstructing CT projection data; do not want to re-run the Reconstruction Project (Task Order 3).

33

Publications/Patents

- Publications permitted and encouraged
- Prior review required by ALERT
  - Process denoted REAP
  - Use “leads to higher false alarms”
  - Do not say “cannot detect”
- May obtain patents on work

34

Additional Information

- Statement of Work (SOW) will be converted to technical specification
- Probably will be insufficient
  - Especially on rules for features and classifier
- Ask questions
  - Answers will be supplied to team
  - Technical spec revised

35

Miscellaneous

- Meeting at ADSA09?
- Meeting at RSNA?
- There will be unfunded participants using the data
- Contract, invoicing issues – contact John Beaty
- Technical issues – contact Castanon, Crawford, Martz

36

Furnished to ATR Developers

- Technical specification
- Database
  - CT images
  - Ground truth labels
  - Sample ATR
  - Tools for assessing detection, PD/PFA
- Access to subject matter experts

37

Performer Future

- Developing vendor-neutral ATRs.
- Working with vendors
- Work on other modalities
- Prediction of detection capability of future scanners

38

Miscellaneous

- Thank you to Steve Azevedo, LLNL, for taking minutes during kickoff meeting
- There will be unfunded participants using the dataset and participating in meetings.

39

THE END

40

11.2.2 Statement of Work

"Statement of Work: Advances in Automatic Target Recognition (ATR) for CT-Based Object Detection System – Task Order 4 & ATR Initiative"

Statement of Work

Advances in Automatic Target Recognition (ATR) for CT-Based Object Detection System – Task Order 4 & ATR Initiative

Version 4

Conducted by DHS Center of Excellence, ALERT at Northeastern University

For

U.S. Department of Homeland Security Directorate of Science and Technology, Explosives Division

1 Executive Summary

The ALERT Center of Excellence at Northeastern University has received funding from the Department of Homeland Security for a project entitled *Advances in Automatic Target Recognition (ATR) for CT-Based Object Detection System.* This project is also known as *Task Order Four* and the *ATR Initiative*. The project will address improving CT-based explosive detection equipment by developing improved ATR algorithms. The project is broken down into a number of tasks, with a principal investigator designated to lead each task. The purpose of this document is to provide the statement of work for each task.

Contents

2 Acronyms and Definitions

2.1 Acronyms

ADSA	Algorithm Development for Security Applications, a series of workshops conducted at NEU
ALERT	Awareness and Localization of Objects-Related Threats, A Department of Homeland Security Center of Excellence at NEU
ATR	Automated target recognition
COE	Center of excellence, a DHS designation
CT	Computerized tomography
DAS	Data acquisition system
DHS	Department of Homeland Security
EDS	Explosives detection system. An EDS is composed of a CT scanner, an ATR algorithm, and a baggage viewing workstation.
FOV	Field of view
HME	Homemade explosive
LLNL	Lawrence Livermore National Laboratory
NEU	Northeastern University
PD	Probability of detection
PFA	Probability of false alarm
PI	Principal investigator; a synonym of researcher
ROC	Receiver operator characteristic
SME	Subject matter expert
SSI	Sensitive security information
SOW	Statement of work
TBD	To be determined
TO1	Task Order 1. The Segmentation Initiative
TO3	Task Order 3. The Reconstruction Initiative
TO4	Task Order 4. This project: the ATR Initiative
TSA	Transportation Security Administration

2.2 Terms

Algorithm	The mathematical steps (or recipe) used to perform a defined problem. This definition does not include computer code.
Corrected data	Raw data (projections) after being corrected for scanner and object imperfections, and the logarithm taken.
Correction	A synonym for pre-processing.
Detection	When an ATR produces an object that matches a target in the ground-truth.
False alarm	When an ATR produces an object that does not match a target in the ground-truth.
Ground-truth	Label images showing the locations of targets in CT scans

Label Image	An image showing to which target or object a pixel belongs. Label images are generated by ATRs or by a program used to create ground-truth
Object	The set of pixels in a CT image associated with a target. The object can be obtained from a program that generates ground-truth or by an ATR.
Phantoms	A numerical description of the contents of a bag. Or, a physical piece of luggage containing known geometric shapes.
Precision	A measure of how well an object generated by an ATR matches the ground-truth.
Pre-processing	The reconstruction step converting raw data to corrected data.
Projection data	Collections of line-integrals of objects.
Raw data	Projection data directly from the x-ray sensor or data acquisition (DAS)
Recall	A measure of how well an object generated by an ATR matches the ground-truth.
Researcher	A performer for tasks described in TO4. A synonym of PI.
Security vendor	A company developing EDS equipment. The equipment may or may not be deployed in the airports in the United States. The list of security vendors includes L-3 Communications, Reveal Detection, Morpho Detection, Analogic, Rapiscan and SureScan.
Targets	Objects that will be evaluated assessing performance of an ATR. DHS uses the definition "Automatic *Threat* Recognition." The term *target* is used herein instead of *threat* in order to emphasize that explosives and explosive simulants will not be used in this project.
Task order	A type of funding vehicle that DHS uses for funding performers.
Third-party	A person or group not working for a security vendor. A third-party works in academia or in industry other than the security vendors

3 Definitions

3.1 CT Image

A CT image corresponds to a set of axial slices resulting from a CT scan of a bag. Each axial slice contains 512x512 pixels. The slices are spaced at approximately 1.5 mm. The number of slices per bag, *N*, depends on the length of the bag. The total number of pixels is *N*x512x512. The slices are single energy.

3.2 Label Image

A label image is a set of slices that indicate if a pixel in a CT image corresponds to an object. The number of pixels in a label image is the same as a CT image, *N*x512x512. The object can be obtained from a program that generates ground-truth or by an ATR. An image pixel can be assigned to only one object. The value of the background is zero. In a label image containing ground-truth the pixel values correspond to physical labels assigned to each target. In a label image generated by an ATR the pixel values are assigned by the ATR and reported in a log file.

3.3 Log File

A log file is created by an ATR and contains the following information.

1. The number of objects found in a CT image

2. For each object,
    a. Mass
    b. Density
    c. Bounding box
    d. Object number in a label image

The log file is human readable.

3.4 ATR

An ATR is a program that executes an algorithm to find objects corresponding to targets. The input to an ATR is a CT image. The outputs of an ATR are a label image and a log file. An ATR may be comprised of the following steps.

1. Segmentation
2. Feature extraction
3. Correction for CT imperfections
4. Classification

3.5 Detection

A detection occurs when an object declared by an ATR *matches* the ground-truth for a target. The term *match* is defined in terms of *recall*, *R*, and *precision, P*. Let *G* correspond to the set of pixels in the ground-truth for a target. Let *S* correspond to the set of pixels declared to be an object by at ATR. Then recall and precision are defined as follows.

$$R = \frac{volume(G \cap S)}{volume(G)}$$

$$P = \frac{volume(G \cap S)}{volume(S)}$$

For this project a detection occurs when $R \geq 0.5$ and $P \geq 0.5$.

Some of the literature cited in Section 8 uses the terms *wholeness* and *exclusiveness*, which correspond to *recall* and *precision*, respectively.

3.6 False Alarm

A false alarm occurs when an ATR creates an object that does not match the requirements for a detection.

3.7 Scoring Program

A scoring program takes as inputs the label image produced by an ATR and the ground-truth label image and determines the number of detections and false alarms. The program generates the following information.

1. The number of detections

2. The number of false alarms
3. For each detection or false alarm
    a. Type (detection or false alarm)
    b. Value of target in the ground-truth
    c. Value of the object in the ATR's label image
    d. Precision, *P*
    e. Recall, *R*
4. The number of targets in the ground-truth that are not detected along with the following information.
    a. Value of target in the ground-truth
    b. For each object declared by the ATR that intersects the target:
        i. Value of the object in the ATR's label image
        ii. Precision, *P*
        iii. Recall, *R*

4 Introduction

The Department of Homeland Security (DHS) has requirements for future explosives detection systems (EDS) that include increased probability of detection and decreased probability of false alarm for a larger set of objects and with reduced minimum masses. The larger set of objects includes certain types of homemade explosives (HME). There are indications that these requirements for future EDS equipment may be difficult to achieve with the technologies presently deployed in the field. In order to resolve these issues, DHS has adopted the strategy of augmenting the capabilities and capacities of the vendors of EDS equipment with the involvement of third parties. Third parties are defined as researchers from academia and industry other than the vendors.

DHS has funded ALERT to execute a project denoted the *Automated Target Recognition (ATR) Initiative*, which is also known as *Task Order Four* (TO4). The goal of this project is to involve third parties in the development of ATR algorithms that could eventually be deployed by the incumbent vendors. The work will be led by the Northeastern University component of the ALERT DHS Center of Excellence (COE). The investigators for the projects will be comprised of researchers both within and outside of the current group of people being supported by the COE.

The research is designed with the following outcomes for DHS.

- The program will improve ATRs. The improved target recognition may lead to decreased minimum target mass, increased target population coverage, increased probability of detection and decreased probability of false alarm.
- The program will increase involvement of third parties via the availability of common CT datasets, and tools, which will increase the work in target recognition, and the number of students who can join the workforce of the vendors and DHS.
- The program will foster collaboration between academics, national laboratory personnel and incumbent security industry vendors.

Technical interchange will be facilitated near the end of the project so that the researchers can present their results to the security vendors, DHS and other third-parties. The results will be documented in a final report for DHS.

The project is broken down into a number of tasks, with a principal investigator designated to lead each task. The purpose of this document is to provide the statement of work for each task.

5 Overview of Project

An overview of the project is described in this section.

1. Bags will be packed to represent what is found in stream of commerce in airports.
2. The targets will be saline, polymer clay and rubber sheets.
3. The bags will be scanned on a medical CT scanner resulting in images of the bags. The scans will be single energy.
4. The contents and their placement in bags will be documented.
5. The voxels corresponding to the targets in the images will be marked and stored in label images. This marking is known as ground truth.
6. ATRs will be developed using the images and the ground truth.
7. The ATRs will be assessed using the following criteria.
    a. Minimizing the probability of false alarm (PFA) for a specified probability of detection (PD)
    b. [Optional] Maximizing the area under the receiver operator characteristic (ROC)
    c. Minimal use of algorithms for specific target configurations (known as corner cases)
    d. Minimal overtraining on test data
    e. Novelty compared to the prior art
    f. Ability to detect targets in difficult configurations
    g. Potential to be extended to detect additional targets
8. Software will be supplied to compare the results of ATRs with the ground-truth.
9. The following items will be supplied to assist the development of ATRs
    a. A benchmark (notional) ATR so that common functions (e.g., reading and writing images and results) do not have to be replicated by each PI. The benchmark ATR may be updated as necessary during the course of the project.
    b. The segmentation algorithms developed for the Segmentation Initiative (Task Order 1) may be available for use in this project. Some of the algorithms may be patented. See Section 8.2 for additional information.
    c. A bibliography describing related prior art in the ATR field. A first version of this bibliography can be found in Section 8.

6 Statement of Work

The project is broken down into a number of tasks. A PI is designated to lead each task. The purpose of this section is to list the statement of work for each task.

There are common requirements for all of the tasks, as listed in Section 7.

The researchers will be assisted by the three subject matter experts (SME) listed in Section 9.4.

6.1 Task 1: Project Plan and Status Reports

6.1.1 Synopsis

The PI shall develop an integrated project plan for the project. The PI will coordinate the development of this plan and will provide engineering and technical review. The PI will coordinate and conduct informal technical workshops with knowledgeable and interested third party algorithm developers as part of the planning process. The PI will prepare status reports for DHS.

6.1.2 Specific Activities

1. Write an integrated project plan for this project.
2. Conduct periodic meetings with all the researchers.
3. Conduct a kick-off meeting with the researchers.
4. Provide status reports to DHS.

6.1.3 Deliverables

1. Integrated project plan
2. Status reports to DHS

6.1.4 Notes

1. This SOW may be the basis of the integrated project plan.

6.2 Task 2: Software Development and Execution

6.2.1 Synopsis

The PI shall develop software to facilitate the development and scoring of ATRs. The software shall be for the following tasks: benchmark ATR, scoring and ground-truth labeling. The PI shall also apply the ground-truth labeling software to the scans of targets on a medical scanner. All the software shall be revised as necessary based on feedback from stakeholders.

6.2.2 Specific Activities

1. Write technical proposal describing how task will be accomplished. Proposal shall be a PowerPoint presentation and will be reviewed by the SMEs.
2. Develop the following software tools based on tools already developed for the Segmentation Initiative (Task Order 1):
    a. Benchmark ATR - see Section 3.4 for additional information
    b. Scoring program – see Section 3.7 for additional information
    c. Ground-truth labeling program – to automatically or semi-automatically identify the pixels corresponding to targets in CT images

3. Validate the reconstructions from the medical scanner by reviewing the images visually
4. Generate ground-truth from scans of targets on a medical CT scanner. A TBD metric will be developed for the accuracy on the label images.
5. Validate the ground-truth by overlaying the ground-truth with the CT images and reviewing the overlaid images visually.

6.2.3 Deliverables

1. Technical proposal
2. Software tools (source code and executable)
    a. Benchmark ATR
    b. Scoring program
    c. Ground-truth labeling program
3. Ground-truth label images from scans of targets on a medical CT scanner
4. Documentation
    a. Algorithm descriptions
    b. Build and execution instructions
5. Source code and *Makefiles*
6. Verification of medical images (an email is sufficient)
7. Verification of ground truth images (an email is sufficient)

6.2.4 Notes

1. The tools shall be developed in the C programming language and run under Linux.
2. The tools should be based on the tools developed by Seemeen Karimi for the Segmentation Initiative. These tools include programs for segmentation, scoring and semi-automatic created of ground-truth.
3. Changes to these tools may include.
    a. Changing the formats of:
        i. Input and output images and labels
        ii. Log files
    b. Separate paths for sheets and bulk targets
    c. Adding a classifier
4. Ground-truth shall be generated for approximately 200 bags. The ground-truth shall be generated for the targets and not for all the objects in a bag.
5. Ground-truth may be created using automatic, semi-automatic or manual methods.
6. Acceptance criteria for the accuracy of the ground-truth data shall be supplied at a later date.
7. All source code shall be put under revision control.

6.3 Task 3a: CT Scanning

6.3.1 Synopsis

The PI shall provide a medical CT scanner on which scans of targets can be obtained. The resulting images shall be provided to the researchers developing ATRs. The raw data shall be saved for future task orders.

6.3.2 Specific Activities

1. Scan bags on a single energy medical CT scanner
2. Reconstruct images
3. Save raw data from the scans

6.3.3 Deliverables

1. Images
2. Raw data
3. Assurance that the scanning is working according to its specification

6.3.4 Notes

1. The bags and their contents will be provided to the PI of this project (see Task 3b for details).
2. Approximately 200 bags shall be scanned.
3. The scan and reconstruction protocols from the Reconstruction Initiative shall be used. The reconstructions shall be using the scanner's reconstruction software.
4. Additional details about this task can be found in following document, which is incorporated by reference: "Scanning Requirement Specification for the ALERT Automated Threat Recognition Initiative (Task Order 4)."

6.4 Task 3b: Bag Packing

6.4.1 Synopsis

The PI shall procure, pack and document the bags and their contents for scanning on the medical CT scanner. The PI shall travel to the site of the medical CT scanner to support the scanning of the bags on the medical scanner.

6.4.2 Specific Activities

1. Procure bags and their contents
2. Mark objects with unique numerical identifier
3. Pack bags
4. Document the contents of bags

6.4.3 Deliverables

1. Packed bags
2. Documentation
    a. Object descriptions including
        i. Numerical identifier
        ii. Mass
        iii. Dimensions
        iv. Photographs
    b. Packing information

6.4.4 Notes

1. Approximately 200 bags shall be scanned.
2. A specification will be supplied by the SMEs for how to pack the bags.

3. The PI for this task shall be supervised by the SMEs.
4. The bags used for the Segmentation and Reconstruction Initiatives may be used in this project.
5. Multiple targets may be present in each bag.
6. Targets shall not be physically touching each other.
7. Targets are contiguous (i.e., not broken into multiple, dis-contiguous components).

6.5 Task 4: Automatic Target Recognition (ATR) Development

6.5.1 Synopsis

The PI shall develop an ATR based on the technical requirements supplied in this document. The technical details may be augmented by an additional technical specification.

6.5.2 Specific Activities

1. Write a technical proposal describing how task will be accomplished. The proposal shall be a PowerPoint presentation and will be reviewed by the SMEs.
2. Develop an ATR to minimize PFA for a TBD PD

6.5.3 [Optionally] Develop methods to maximize the area under an ROCDeliverables

1. PD/PFA results
2. [Optional] ROC
3. Detection results for all targets
4. Label images for all scans of bags
5. Technical proposal

6.5.4 Notes

1. Multiple researchers will be developing ATRs. The researchers shall develop their algorithms independently.
2. The ATRs may use the benchmark ATR program for reading and writing images and reporting results. The benchmark ATR may be revised based on requests from the researchers.
3. The input to the ATR shall be images. Raw data shall not be used by the ATR.
4. There shall not be an upper limit on mass or volume.
5. Detection should be independent of shape, size, location, orientation, clutter, and concealment. This means that the researcher should not try to meet the PFA requirement by not detecting configurations of targets that lead to high false alarms.
6. PD may be weighted to emphasize targets whose images are corrupted by CT artifacts.
7. The researchers are requested to:
    a. Separate the data into training and test sets
    b. Not over-train on the data
    c. Design their ATRs to be extensible so that additional targets can be considered in the future.

8. Some scans will be supplied to the researchers close to the end of the project in order to provide additional testing of their ATRs.
9. Containers (e.g., bottles) for liquids are not considered to be part of the target.
10. The ATRs shall be different than the methods presented in the prior art.
11. The format of reporting the results of an ATR will be supplied later.
12. There is no requirement to report the type of target (e.g., saline, modeling clay or rubber sheet).

6.6 Task 8: Technical Interchange[1]

6.6.1 Synopsis

The PI shall coordinate and conduct a meeting for the demonstration of the project research results. The following groups will be invited to the meeting: vendors of security equipment, government agencies and national labs.

6.6.2 Specific Activities

1. Coordinate and moderate technical interchange.

6.6.3 Deliverables

1. Presentations
2. Minutes

6.7 Task 9: Final report

6.7.1 Synopsis

The PI will write a final report for the project.

6.7.2 Specific Activities

1. Write final report.

6.7.3 Deliverables

1. Final report

6.7.4 Notes

1. Two versions of the final report may be required: one for DHS and for distribution in the public domain.

7 Common Requirements

7.1 Targets

7.1.1 Materials

1. Saline doped to have a densities overlapping with other liquids commonly found in bags.

[1] The gap in numbering the tasks is due to matching the number of tasks in DHS's contract with ALERT for this project.

2. Modeling clay (polymer) doped with TBD material to vary its linear attenuation coefficient
3. Rubber sheets

7.1.2 Requirements

1. Minimum mass: 250 g
2. Maximum mass: None
3. Minimum thickness for sheets: TBD

7.2 Common Supplied Materials

1. Images from scans of bags on medical CT scanner
2. Ground-truth label images
3. File formats
4. Description of all scanned objects
5. Scanner parameters
6. Description of vendor reconstruction
7. Benchmark ATR
8. Scoring software

7.3 Common Activities

All of the researchers are required to perform the following activities.

1. Write a project plan addressing the following topics.
    a. Objectives
    b. Scope
    c. Deliverables
    d. Acceptance criteria
    e. Methods
    f. Schedule
2. Revise the project plans based on feedback provided by the SMEs.
3. Participate in a project kickoff meeting via web conferencing
4. Participate in monthly teleconferences
5. Provide monthly written status reports.
6. Review the plan for collecting CT data
7. Present results at a technical interchange (travel expense paid by researcher) addressing the following topics.
    a. Objectives
    b. Results
    c. Methods
    d. Computational expense and potential for reduction/acceleration
    e. Recommendations for future work.
8. Write a final report addressing the following topics.
    a. Objectives
    b. Results

    c. Methods
    d. Computational expense and potential for reduction/acceleration
    e. Recommendations for future work
    f. Biographies of members of research team
    g. Bibliography
9. Host visits by the SMEs
10. Collaborate with other researchers as necessary

7.4 Common Technical Details

7.4.1 Data Formats

1. Images: DICOM
2. Labels: FITS

7.5 Common Notes

1. There are no requirements on the execution speed for algorithms.
2. Researchers shall describe their computational environment and the execution times of their algorithms.
3. An online repository will be supplied for this project that includes the following information.
    a. Documents
    b. Schedules
    c. Data
        i. CT images
        ii. Ground-truth
    d. Mailing lists
    e. Software
        i. Benchmark ATR
        ii. Scoring software

8 Bibliography of Related Information

8.1 ADSA Workshops

ATR has been discussed at all of the ADSA workshops. The most relevant material is as follows.

1. ADSA02
    a. Martz's presentation on segmentation challenges
    b. Martz's presentation on LLNL's 3rd party algorithm development project
    c. Crawford and Martz's overview of EDS equipment; see Section 8.3.2.
    d. Crawford's review of automated target detection algorithms
2. ADSA08
    a. This workshop dealt almost entirely with ATR

The final reports for the ADSA workshops can be found at the following link.

https://myfiles.neu.edu/groups/ALERT/strategic_studies/

8.2 Segmentation Initiative

The goal of Segmentation Initiative was to have five research groups develop algorithms for segmenting objects from volumetric CT data. These results may be used in the present project. In addition, the researchers in the present project may collaborate with the researchers in the Segmentation Initiative. Additional details can be found in the final report for the Segmentation Initiative at the following link:

https://myfiles.neu.edu/groups/ALERT/strategic_studies/SegmentationInitiativeFinalReport.pdf

8.3 Literature

1. Merzbacher and Gable, “Applying Data Mining To False Alarm Reduction In An Aviation Explosives Detection System.”
2. Crawford and Martz, “Overview of Deployed EDS Technologies,” LLNL Technical Report, LLNL-TR-417232, September 24, 2009.
3. Ying, Z., Naidu, R., and Crawford, C., "Dual Energy Computed Tomography for Explosive Detection," Journal of X-ray Science and Technology, vol. 14, 2006, pp. 235-256.

8.4 United States Patents

A partial list of patents related to this project can be found in the following sections. The researchers are encouraged to read these patents. In addition, online databases (e.g., Google Patents or the US Patent Office) will show other related patents by checking the citations in these patents or by searching for other patents that cite the listed patents below.

8.4.1 Morpho Detection

1. 8260020, Garms
2. 8090169, Ioannou
3. 8090150, Garms
4. 7840030, Schmeigel
5. 7839971, Bendahan
6. 7660457, Schmeigel
7. 7889836, Kaucic
8. 2010/0284618 A1, Ingerman
9. 2009/0169079, Garms
10. 2009/0087012, Merzbacher
11. 2008/0253653, Gable
12. 2008/0169104, Ioannou
13. 2007/000863851, Merzbacher

8.4.2 L-3 Communications

1. 5712926, Eberhard
2. 5905806, Eberhard
3. 7613316, Mahdavieh

8.4.3 Analogic

1. 5,901,198, Ruth
2. 6,111,974, Hiraoglu
3. 6,075,871, Simanovsky
4. 6,128,365, Bechwati
5. 6,026,143, Simanovsky
6. 6,076,400, Bechwati
7. 6,108,396, Bechwati
8. 6,078,642, Simanovsky
9. 6,026,171, Hiraoglu
10. 6,272,230, Hiraoglu
11. 6,035,014, Hiraoglu,
12. 6,317,509, Simanovsky
13. 6,067,366, Simanovsky
14. 6,195,444, Simanovsky
15. 6,345,113, Crawford
16. 7,277,577, Ying
17. 7,190,757, Ying
18. 7,801,348, Ying

9 Technical Leadership and Program Management

9.1 Principal Investigator

The PI for this project will be Michael Silevitch, NEU, and ALERT co-Director.

9.2 Program Management Lead

The program management lead will be John Beaty, Director of Technology Development for ALERT.

9.3 Technical Lead

The technical lead for this project will be Carl Crawford, Csuptwo, LLC.

9.4 Subject Matter Experts (SME)

The leadership team will be assisted by the following SMEs.

1. Carl Crawford, Csuptwo, LLC
2. Harry Martz, Lawrence Livermore National Laboratory
3. David Castanon, Boston University

The SMEs will seek advice and guidance from the incumbent vendors.

10 Schedule

The following table shows approximate dates for the key milestones in the project.

Milestone	Owner(s)	Month
Project start	PI for Task 1	1
Kickoff meeting	PI for Task 1	1
Scanning on medical scanner and image release	PIs for Tasks 3a and 3b	1
Technical proposal from researchers	PIs for Tasks 2 and 4	1
Technical description of project	PI for Task 1	1
Initial release of benchmark ATR and scoring software	PI for Task 2	1
Ground truth	PI for Task 2	2
Second release of benchmark ATR and scoring software	PI for Task 2	2
Test data supplied to researchers	PI for Task 2	16
Technical interchange	PI for Task 8	17
Final reports due from researchers	PIs for Tasks 2 and 4	17
Final report for DHS	PIs for Tasks 2 and 4	18
Project end	PI for Task 1	18
Status reports and teleconferences	PIs - all	Monthly
Technical discussions with SMEs	PIs - all	Quarterly

11 Contractual

11.1 Contract

Researchers will have to sign a contract that stipulates the following.

1. CT data cannot be distributed to other parties.
2. Resulting intellectual property is owned by researcher
3. Publication of results is permissible after review by ALERT.
4. Publications may not discuss the following topics
    a. Connection to actual explosives
    b. Difficulties in achieving PD requirements
    c. SSI information
    d. Classified information

11.2 Funding types

Researchers may receive funding using the following mechanisms.

1. Grant through an academic institution. This is a fixed price contract. Institutional overhead will not be covered by ALERT/NEU.
2. Consultant to NEU
3. Purchase order to a commercial company

12 Revision History

Version	Changes
1	First release.
2	Revised based on feedback from stakeholders.
3	Revised based on feedback from two possible researchers and feedback from Jeff Kallman, LLNL
4	Corrected formatting bug due to Word's tracking feature

11.2.3 Bibliography

"ATR Initiative Annotated Bibliography"

ATR Initiative Annotated Bibliography

1 ADSA Workshops

ATR has been discussed at all of the ADSA workshops. The most relevant material is as follows.

1. ADSA01
   a. final report
2. ADSA02
   a. Martz's presentation on segmentation challenges
   b. Martz's presentation on LLNL's 3rd party algorithm development project
   c. Crawford and Martz's overview of EDS equipment; see Section 3.4.2.
   d. Crawford's review of automated target detection algorithms
   e. Final report
3. ADSA08
   a. This workshop dealt almost entirely with ATR
   b. Final report

The final reports for the ADSA workshops can be found at the following link.

https://myfiles.neu.edu/groups/ALERT/strategic_studies/

2 Segmentation Initiative

The goal of Segmentation Initiative was to have five research groups develop algorithms for segmenting objects from volumetric CT data. These results may be used in the present project. In addition, the researchers in the present project may collaborate with the researchers in the Segmentation Initiative. Additional details can be found in the final report for the Segmentation Initiative at the following link:

https://myfiles.neu.edu/groups/ALERT/strategic_studies/SegmentationInitiativeFinalReport.pdf

3 United States Patents

A partial list of patents related to this project can be found in the following sections. The researchers are encouraged to read these patents. In addition, online databases (e.g., Google Patents or the US Patent Office) will show other related patents by checking the citations in these patents or by searching for other patents that cite the listed patents below.

3.1 Morpho Detection

1. 8260020, Garms
2. 8090169, Ioannou
3. 8090150, Garms
4. 7840030, Schmeigel

ATR Initiative Annotated Bibliography2

5. 7839971, Bendahan
6. 7660457, Schmeigel
7. 7889835, Kaucic
8. 2010/0284618 A1, Ioannou
9. 2009/0087012, Merzbacher
10. 2008/0253653, Gable
11. 8254676, Ioannou
12. 2009/0226032A1 Merzbacher

~~13.~~ 7,031430, Kaucic
~~14.~~ 7,869,566, Edic
~~15.~~ 8,180,138, Basu

3.2 L-3 Communications

1. 5712926, Eberhard
2. 5905806, Eberhard
3. 7613316, Mahdavieh

3.3 Analogic

1. 5,901,198, Ruth
2. 6,111,974, Hiraoglu
3. 6,075,871, Simanovsky
4. 6,128,365, Bechwati
5. 6,026,143, Simanovsky
6. 6,076,400, Bechwati
7. 6,108,396, Bechwati
8. 6,078,642, Simanovsky
9. 6,026,171, Hiraoglu
10. 6,272,230, Hiraoglu
11. 6,035,014, Hiraoglu,
12. 6,317,509, Simanovsky
13. 6,067,366, Simanovsky
14. 6,195,444, Simanovsky
15. 6,345,113, Crawford
16. 7,277,577, Ying
17. 7,190,757, Ying
18. 7,801,348, Ying
19. 7,302,083, Larson
20. 7,327,853, Ying
21. 7,474,786, Naidu
22. 7,539,337, Simanovsky

ATR Initiative Annotated Bibliography3

8,391,600 Litvin

3.4 Rapiscan

1. 7,876,879, Morton
2. 8,223,919, Morton
3. 8,428,217, Peschmann

4 Literature

4.1 Miscellaneous

1. Merzbacher and Gable, "Applying Data Mining To False Alarm Reduction In An Aviation Explosives Detection System."
2. Crawford and Martz, "Overview of Deployed EDS Technologies," LLNL Technical Report, LLNL-TR-417232, September 24, 2009.
3. Ying, Z., Naidu, R., and Crawford, C., "Dual Energy Computed Tomography for Explosive Detection," Journal of X-ray Science and Technology, vol. 14, 2006, pp. 235-256.
4. R.C. Smith and J.M. Connelly, CT TECHNOLOGIES, in Oxley et al.
5. Hyland, S., Bitner, C., Cooks, G., Crawford, C., Garrick, J., Gatsonis, C., Glover, G., Lele, S., Martz, H., and Meeker, W., "Engineering Aviation Security Environments - Reduction of False Alarms in Computed Tomography- Based Screening of Checked Baggage," Commitee on Engineering Aviation Security Environments - False Positives from Explosive Detection Systems, National Research Council of the National Academies, July 19, 2013. The final report is available from www.nap.edu/catalog.php?record_id=13171.

4.2 Siemens from Segmentation Initiative

1) Leo Grady, Vivek Singh, Timo Kohlberger, Christopher Alvino and Claus Bahlmann, "Automatic Segmentation of Unknown Objects, with Application to Baggage Security", Proc. of ECCV, pp. 430-444, 2012.

2) Wei Li, Gianluca Paladini, Leo Grady, Timo Kohlberger, Vivek Singh, Claus Bahlmann, "Luggage Visualization and Virtual Unpacking", Workshop at SIGGRAPH Asia, 2012.

1) Timo Kohlberger, Vivek Singh, Chris Alvino, Claus Bahlmann and Leo Grady, "Evaluating Segmentation Error Without Ground Truth", Proc. of MICCAI, Vol. 7510, pp. 528-536, 2012.

4.3 University of East Anglia (UEA) from Segmentation Initiative

[1] Bangham, R. Harvey, P. Ling, and R. Aldridge. Morphological scale-space preserving transforms in many dimensions. Journal of Electronic Imaging, 5:283–299, 1996.

[2] J. Bangham, P. Ling, and R. Young. Multiscale recursive medians, scale-space, and transforms with applications to image-processing. IEEE Trans Image Processing, 5(6):1043–1048, June 1996.

ATR Initiative Annotated Bibliography4

[3] J. A. Bangham, P. Chardaire, C. J. Pye, and P. D. Ling. Multiscale nonlinear decomposition: The sieve decomposition theorem. IEEE Transactions on Pattern Analysis and Machine Intelligence, 18(5):529–539, May 1996.

[4] J. Bangham, J. Hidalgo, R. Harvey, and G. Cawley. The segmentation of images via scale-space trees. In British Machine Vision Conference, pages 33–43, 1998.

[5] P.Southam and R.Harvey. Texture classification via morphological scale-space: Tex-Mex features. Journal of Electronic Imaging, Vol. 18, 043007 Nov 2009

4.4 Stratovan from Segmentation Initiative

David F. Wiley, Deboshmita Ghosh, and Christian Woodhouse. "Automatic Segmentation of CT scans of Checked Baggage", Proceedings of the 2nd International Meeting on Image Formation in X-ray CT, Salt Lake City, Utah, pp. 310-313, June 2012.
http://www.stratovan.com/about/pdf/AutomaticSegmentationOfCtScansOfCheckedBaggage.pdf

David F. Wiley. "Analysis of Anatomic Regions Delineated from Image Data", US Patent 8,194,964 , Filing data April 27, 2009, Issue date June 5, 2012.

4.5 TeleSecurity Sciences from Segmentation Initiative

[1] J. Kwon, J. Lee, S. Song, Extraction of objects from ct images by sequential segmentation and carving, US Patent Pending, Filing date Dec. 12, 2012, Pub. No. US20130170723 A1

4.6 Other Research Papers

4.6.1 Cranfield U. (UK)

[1] N. Megherbi, G. Fitton, and T. Breckon, "A classifier based approach for the detection of potential threats in CT based baggage screening", *in Proc, 2010 IEEE ICIP*, pp. 1833-1836.
[2] G. Fitton, T. Breckon, and N. Megherbi, "Object recognition using 3D SIFT in complex CT volumes", in *Proc. British Machine Vision Conference*, pp. 1-12. 2010.
[3] G. Fitton, T. Breckon, and N. Megherbi, "A comparison of 3D interest point descriptors with application to airport baggage object detection in complex CT imagery", *Pattern Recognition*, vol. 29, no. 9, pp. 2420-2436, 2013.

4.6.2 Tsinghua U. (China)

[4] W. Bi, Z. Chen, L. Zhang, and Y. Xing, "A volumetric object detection framework with dual-energy CT", in Proc. 2008 Nuclear Science Symposium, pp. 1289-1291, 2008.

5 Revision History

Version	Date	Author	Revisions
0	7/16/2013	Crawford	Initial release.
1	9/23/2013	Crawford	Additions provided by Sam Song (Telesecurity Sciences) and Sergey Simanovsky (Analogic)

11.3 Scanning

11.3.1 Scanner Requirement Specification

"Scanning Requirement Specification for the ALERT Automated Threat Recognition Initiative (Task Order 4)"

Scanning Requirement Specification for the ALERT Automated Threat Recognition Initiative (Task Order 4)

Version 2

TO4 Scanning Requirement Specification, Page 2

Contents

TO4 Scanning Requirement Specification, Page 3

1 Introduction

1.1 Purpose

A specification is provided for scanning objects on a CT scanner to support ALERT's Automated Threat Recognition (ATR) Initiative, which is also known as Task Order 4 (TO4). The purpose of TO4 is to develop ATR algorithms for CT-based explosive detection equipment.

1.2 Scope

Guidelines are provided to achieve the following purposes.

1. Identify CT scanners that can be used to perform the scanning described herein.
2. Negotiate with vendors of said CT scanners to determine the cost of scanning and to determine recommended scanning technique.
3. Scanning objects.
4. Distributing the results of scanning.

1.3 Acronyms

ALERT	Awareness and Localization of Objects-Related Threats, A Department of Homeland Security Center of Excellence at NEU
ATR	Automated threat recognition
COE	Center of excellence, a DHS designation
CT	Computerized tomography
DAS	Data acquisition system
DHS	Department of Homeland Security
EDS	Explosives detection system. An EDS is composed of a CT scanner, an ATR algorithm, and a baggage viewing workstation.
FOV	Field of view
HME	Homemade explosive
NEU	Northeastern University
SSI	Sensitive security information
TBD	To be determined
TSA	Transportation Security Administration

1.4 Definitions

Algorithm	The mathematical steps (or recipe) used to perform a defined problem. This definition does not include computer code.
Corrected data	Raw data (projections) after being corrected for scanner and object imperfections, and the logarithm taken.
Correction	A synonym for pre-processing.

Security vendor	A company developing EDS equipment. The equipment may or may not be deployed in the airports in the United States. The list of security vendors includes L-3 Communications, Reveal Detection, Morpho Detection, Analogic, Rapiscan and SureScan.
Object of interest	Objects that will be evaluated assessing performance of an ATR
Phantoms	A numerical description of the contents of a bag. Or, a physical piece of luggage containing known geometric shapes.
Pre-processing	The reconstruction step converting raw data to corrected data.
Projection data	Collections of line-integrals of objects.
Raw data	Projection data directly from the x-ray sensor or DAS.
Researcher	A performer for tasks described in TO4.
Task order	A type of funding vehicle that DHS uses for fund performers.
Third-party	A person or group not working for a security vendor. A third-party works in academia or in industry other than the security vendors
Vendor	The provider of a CT scanner for this project.

1.5 Requirement Levels

Shall The word *shall* denotes what is required.

Should The word *should* denotes what is desirable but not required.

May The word *may* denotes allowable but not required behavior.

2 Background

DHS has requirements for future explosives detection systems (EDS) that include increased probability of detection and decreased probability of false alarm for a larger set of objects and with reduced minimum masses. The larger set of objects includes certain types of homemade explosives (HME). There are indications that these requirements for future EDS equipment may be difficult to achieve with the technologies presently deployed in the field. In order to resolve these issues, DHS has adopted the strategy of augmenting the capabilities and capacities of the vendors of EDS equipment with the involvement of third parties. Third parties are defined as researchers from academia and industry other than the vendors.

DHS has funded ALERT to execute a project denoted the Automated Threat Recognition (ATR) Initiative, which is also known as Task Order 4 (TO4). The goal of this project is to involve third parties in the development of ATR algorithms that could eventually be deployed by the incumbent vendors. The work will be led by the Northeastern University component of the ALERT DHS COE. The investigators for the projects will be comprised of researchers both within and outside of the current group of people being supported by the COE.

The research is designed with the following outcomes for DHS.

- The program will improve ATRs. The improved threat recognition may lead to decreased minimum threat weight, increased threat population coverage, increased probability of detection and decreased probability of false alarm.
- The program will increase involvement of third parties via the availability of common CT datasets, and tools, which will increase the work in threat recognition, and the number of students who can join the workforce of the vendors and DHS
- The program will foster closer collaboration between academics, national laboratory personnel and incumbent security industry vendors.

Technical interchange will be facilitated near the end of the project so that the researchers can present their results to the security vendors, DHS and other third-parties.

The task order effort and results will be documented in a final report for DHS.

The purpose of this document is to provide the requirements for a CT scanner that can be used to collect the data that will be used to develop and evaluate the ATRs resulting from this project.

3 Scanner and Scanning Requirements

3.1 Scanner Requirements

Parameter	Requirement	Comments
Spatial resolution	< 1 mm, isotropic	Resolution is 10% of modulation transfer function (MTF)
Potential	≥ 140 kV peak	
Current	≥ 100 mAs	
Bowtie filter	None	
FOV, scan and recon	≥ 50 cm	
Scan length	≥ 50 cm	
Acquisition mode	Helical	Single detector row sufficient
Spectral CT	Single energy is sufficient	Dual energy scanning is desirable
Image quality	Consistent with third-generation commercial CT scanners being marketed/sold today	
Raw data availability	Yes	
Corrected data availability	Yes. Only apply following steps: 1. Air 2. Offset 3. Underflow clamping 4. Logarithm 5. Bad detector 6. Missing view	

TO4 Scanning Requirement Specification, Page 6

3.2 Scanning Requirements

Parameter	Requirement	Comments
Total number of scans	200	
Objects if interest	Saline, modeling clay, rubber sheets	Explosives and their simulants will not be scanned.

3.3 Labor and Distribution Requirements

Parameter	Requirement	Comments
Who controls what is scanned?	ALERT	Vendor does not control what is scanned.
Who controls how scans are taken?	ALERT	Vendor does not control how scans are taken.
Who controls who can use the data (projections and meta data)?	ALERT	Vendor does not control who can use the data (projections and meta data).
Who is charge of reviewing manuscripts before publication?	ALERT	Vendor does not have right to review publications. Vendor can provide standard language for describing the scanner and acknowledging the vendor.
Vendor preferential access to results	No	Vendor will not be given right to see results before anyone else.
Who controls how data is distributed?	ALERT	Vendor does not control how data is distributed. Distribution means for example on the internet or with passwords.
Vendor supplied documentation and code to consume the data	1. Scanner description 2. File decoders 3. Offline reconstruction code (source code, executable, and description of build environment) 4. Description of reconstruction and calibration 5. Image quality metrics such as MTF, SSP, and quantum noise 6. Corrected data measurements such as mean and variance of offsets and air 7. Scans of calibration and image quality phantoms	
Vendor supplied reconstructions (images) for all scans	Yes	

TO4 Scanning Requirement Specification, Page 7

Parameter	Requirement	Comments
Vendor supplies representative to discuss scanner	Yes	
Scanner operator (person) supplied by vendor	Yes	
Supplier of objects to be scanned	ALERT	
Retrospective resolution and noise adjustment to match security scanners	May be performed by ALERT	ALERT may choose to blur the projection data or add noise to match the resolution and noise of security scanners.
Classification requirements	All data, documentation and code will be placed into the public domain	Data, documentation and code shall not be SSI, Classified or proprietary

4 Revision History

4.1 Version 1

First release.

4.2 Version 2

Revised based on Song's feedback,.

11.3.2 Scan Plan

"TO4 (ATR Initiative) Scan Plan"

TO4 (ATR Initiative) Scan Plan

Version 7

TO4 Scan Plan, Page 2

Contents

TO4 Scan Plan, Page 3

TO4 Scan Plan, Page 4

1 Introduction

The purpose of this document is to provide the details of the scanning of bags for the ATR Project.

2 Scanning Session Details

2.1 Location

Heartscan
389 Oyster Point Blvd., Suite #3
South San Francisco, CA 94080
Phone: (800) 469-4327

2.2 Shipping Address

Ken Charles
c/o Heartscan
389 Oyster Point Blvd., Suite #3
South San Francisco, CA 94080
Phone: (800) 469-4327

2.3 Times

1. Monday, September 30th - Thursday, October, 4th
2. Start times: 7:30 AM
3. End times: TBD

2.4 Personnel

Name	Affiliation	Roles
Carl Crawford	Csuptwo	Technical lead, offline reconstruction, quality assurance
Alyssa White	ALERT	Object acquisition, pack packing, documentation
Rick Moore	ALERT	Object acquisition, pack packing, documentation
Doug Boyd	Tele-Security Sciences	Interface to Heartscan, project advisor
Sam Song	Tele-Security Sciences	Interface to Heartscan, project advisor
Ken Charles	Tele-Security Sciences	Scanner operator
Harry Martz	LLNL	Subject matter expert
Jeff Kallman	LLNL	Subject matter expert

TO4 Scan Plan, Page 5

3 Definitions

Term/Acronym	Definition
Target	Something that an ATR has to detect
Non-target	Something that an ATR should not detect
Pseudo-target / PT	A target material with sub-minimum mass or a another material with density less than water. A pseudo-target is also a non-target.
Bag	Something to contain targets, non-targets and pseudo-targets
Object	The union of targets, non-targets and bags.

4 Assumptions

1. 150 – 200 bags can be scanned

5 General Requirements

5.1 Tolerances

Tolerances for making test objects listed below shall be as follows.

1. Mass: ±2%
2. Linear dimension: ±2%
3. Volume: ±2%
4. Density: ±2%

5.2 Axes for CT scanner

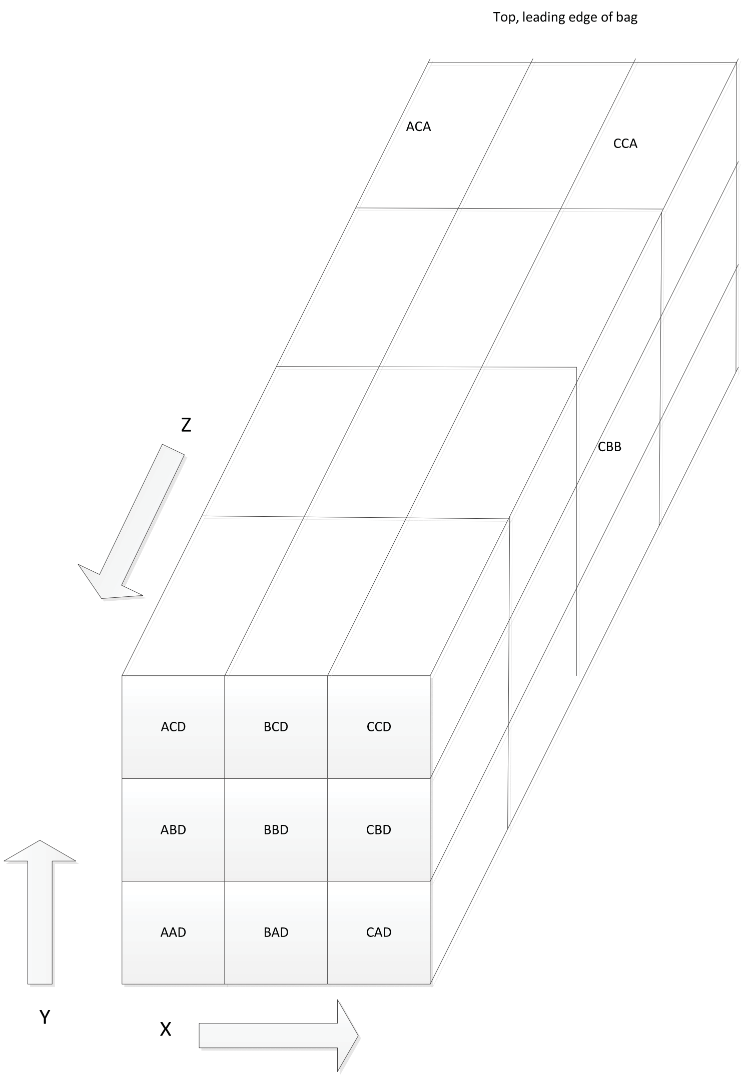
The following axes shall be used for the CT-scanner

1. x: horizontal axis of axial slice
2. y: vertical axis of axial slice
3. z: parallel to direction of table movement for helical scans

5.3 Location Code for Objects Placed in a Bag

A three-letter code is used to note where objects are placed in a bag. The code is of the form xyz, where x, y, and z are letters showing the location along the x, y and z, axes, respectively. The x- and y-axes are split into three sections denoted A, B, and C. The Z-axis is split into four sections denoted A, B, C and D. The following diagram shows some are the location codes map to a bag.

TO4 Scan Plan, Page 6



These codes correspond to the Imatron images as follows.

1. X: A refers to left side of image
2. Y: A refers to bottom of image
3. Z: A refers to slice 1

5.4 Preferred Axes for Objects

1. Cylinders: axis of rotation
2. Sheets: Parallel to conveyor belt
3. Cuboids: Longest dimension

Notes:

1. The preferred axis for an object should be marked on an object.

TO4 Scan Plan, Page 7

5.5 Orientation Codes for Objects

The orientation code is used to specify how the preferred axis of an object is oriented in bag. The values of the code are as follows.

1. Aligned to an axis:
    a. X: aligned to x axis
    b. Y: aligned to y axis
    c. Z: aligned to z axis
2. Not aligned to an axis but in a plane aligned with two axes:
    a. XY: in xy-plane
    b. YZ: in yz-plane
    c. XZ: in xz-plane
3. Other
    a. N: not aligned with an axis and not in plane aligned with two axes

Notes

1. A plus sign (+) sign or minus sign (-) shall be appended to all the orientation codes to show how the preferred axis of an object.

5.6 IDs

1. IDs for targets have to be unique and numeric
2. Each packing or shape of an object has own ID. For example,
    a. Each bottle of saline has its own ID
    b. Each shape (cutting) of a rubber sheet has its own ID
3. The bulk (source) material(s) for targets should also have unique IDs. For example, the box of modeling clay should be given an ID. Each time a piece of clay is cut from the bulk or a piece is molded, it should be given a new ID.

5.7 File Naming Conventions

1. Filenames:
    a. Compatible with Linux and Windows
    b. No spaces
    c. Less than 10 characters
2. Scans should have a unique serial number
3. Filenames should have a specified prefix.
    a. R – raw projection data in DAS counts
    b. S – sinogram (corrected) data
    c. C –CT image data
    d. G – ground truth labels
    e. A – ATR label images

5.8 FTP Site Contents

1. Raw data

TO4 Scan Plan, Page 8

2. Image data
    a. DICOM
    b. FITS
3. Ground truth labels
4. Sinograms
5. Documentation
    a. Spread sheets
        i. Objects
        ii. Packing
    b. Scanning narrative
6. Tools
    a. Franco's code and documentation including
        i. Gregor's FITS's code
        ii. Crawford's parse library

Notes:

1. CT images and label images may be compressed with gzip or zip

6 Objects for Scanning

6.1 General

1. Types
    a. Targets (objects of interest for ATR)
    b. Non-targets
        i. Common items found in checked and carry-on baggage
            1. Produce
            2. Perishables
            3. Fragile items
        ii. Items to contain/conceal targets
    c. Bags
        i. Hard shell
        ii. Soft shell (e.g., duffle bags)
        iii. Plastic bins
2. Sources
    a. Task Order 1 – Segmentation Initiative
        i. Sourced by ALERT
        ii. Presently at Heartscan
    b. Task Order 3 – Reconstruction Initiative

TO4 Scan Plan, Page 9

6.2 Targets

6.2.1 Materials

1. Saline doped to have a densities overlapping with other liquids commonly found in bags.
   a. Range: 1050 MHU – 1150 MHU in steps of 20 MHU
   b. Add 5 g of salt tablet for every 100 mL, which should increase the density by 5%.
   c. The saline should be mixed at a high concentration and then diluted to the appropriate concentration. If you want to get 10 grams of salt per 100 grams of solution, try to dissolve the 10 grams of salt in 70 ml of water and then dilute to get to 100 grams of solution.
2. Modeling clay (polymer) doped with TBD material to vary its linear attenuation coefficient. See following sites for additional information.
   a. http://en.wikipedia.org/wiki/Polymer_clay
   b. www.sculpey.com/products/clays/original-sculpey
3. Rubber sheets and rubber blocks

6.2.2 Requirements

1. Minimum mass: 250 g
2. Maximum mass: None
3. Minimum thickness for sheets: 0.25 inch
4. Maximum thickness for sheets: 1.5 inch
5. Sheet shape:
   a. Not all rectangular
   b. Some with holes cut in the middle

6.3 Pseudo-Targets

1. Pseudo-targets (PT) are one of the following two types
   a. Target materials listed in Section 6.2 with masses ≥ 125 g and < 250 g
   b. Powders with masses greater ≥ 125 g and density < 1 g/cc
2. ATRs are not required to detect PTs. A detection on a PT will be considered to be a false alarm.
3. If possible, ground truth should be created for PTs in the future for future ATR development
4. ≤ 5 PTs may be placed in each bag

6.4 Non-Targets

1. Common objects found in checked and carry-on luggage
   a. Food
   b. Clothing
   c. Cosmetics
   d. Electronics
      i. iPad or equivalent
2. Containment items
   a. Bags
   b. Bins

TO4 Scan Plan, Page 10

3. Clutter items
    a. Electronics
    b. Metal
    c. Large fluid

6.5 Notes

1. All objects have to fit into scan FOV (475 mm)
2. Objects should be less than 840 mm (33 inches)
3. Containers are not part of targets. For example, the bottle containing is not part of the target; only the saline is the target.
4. Targets are contiguous, where the definition of contiguity is TBD

7 Packing

1. Three targets on average per bag
2. Three pseudo-targets on average per bag
3. Targets
    a. Different shapes
        i. Clay: via molding
        ii. Saline: via different containers
        iii. Sheets:
            1. via cutting and bending
            2. Thickness controlled by procurement
    b. Different masses:
        i. x1 – x2 minimum mass
    c. Orientations
        i. Easy
        ii. Medium
        iii. Difficult
    d. Locations
        i. Center
        ii. Periphery
    e. Clutter
        i. Low
        ii. Medium
        iii. Heavy
    f. Emulate some items (e.g., iPad) divested at the check-point
4. Notes
    a. Targets should be shaped differently than non-targets in a bag so that the targets can be visually identified in the CT images. For example, a liquid target should be placed in a round bottle and all non-targets liquids should be placed in rectangular bottles.
    b. TBD if targets can be touching

TO4 Scan Plan, Page 11

c. Non-targets may be repacked. For example:
    i. Coke poured in 250 ml bottle
    ii. Honey put into a plastic bag
    iii. Water
d. Bags shall be set off at least 5 cm from the scanning table with foam
e. Empty spaces in bags may be filled with foam or clothing
f. Object may be held into position using masking tape
g. Containers should not always be completely filled.
h. Simulated detonators, timing devices, power supplies and wires not required
i. All items packed in the bag shall be labeled with an ID. The exception is that clothing and foam does not have to be labeled.

8 Scanning

8.1 Overview

1. Same scanning protocol as TO3 with the exception that only high energy (130 kV) data will be acquired
2. Save raw data
3. Images reconstructed offline

8.2 Scanning Protocol

1. Collimator: 1.5 mm
2. Slice thickness: 1.5 mm
3. kV: 130
4. mA: 630
5. Exposure: 0.1 s for 130 kV
6. Pitch: 1
7. Scan time: 0.100 s
8. Recon Method: Offline
9. Reconstruction kernel: Highest resolution
10. Field of View: 475 mm
11. Tilt: 0.0 deg
12. Slew: 0.0 deg
13. Maximum number of slices: 280

Notes:

1. The maximum number of slices that can be collected with the above protocol is 280 slices, which corresponds to 420 mm or 16.54”. Therefore, bags longer than this length will have to be scanned multiple times at different table locations.
2. Bags should be marked with a leading edge and always scanned in the same orientation.

TO4 Scan Plan, Page 12

9 Reconstruction

1. Images created with xrec offline
2. Limor Martin's protocols
3. DICOM images created
4. Corrected projection data (sinograms) created
5. The center of gravity of each target (in slice) coordinates determined

10 Documentation

10.1 Object Database

1. Per object (target, non-target, bag)
    a. Name
    b. Description (text)
    c. Mass
    d. Dimensions
    e. Photos
        i. Object shown with ruler
        ii. At least one photo shall show the object's ID
    f. ID
2. Notes
    a. Each instantiation of a target (e.g., saline) shall have its own ID

10.2 Packing Database

1. Per bag
    a. ID of bag
    b. Raw data filename(s)
    c. Image file name(s)
    d. Textual description of bag
    e. IDs of non-targets, targets, and pseudo-targets
    f. Orientation of bag when scanned – not required if the bag is aligned with the patient table and the top-leading edge is scanned first and at the top.
    g. Per each target
        i. ID
        ii. Type: saline, sheet,
        iii. Location code of target in the bag; see Section 5.3.
        iv. Orientation code; see Section 5.5.
        v. Description of objects and IDs into which target is inserted
        vi. Code for clutter: light, medium, heavy (code TB(
        vii. Description of clutter
            1. L: low
            2. M: medium

TO4 Scan Plan, Page 13

3. H: heavy

viii. Bounding box for target (in slice) coordinates

h. Unpacking video (TBD)

i. Unpacking photos

2. A log (i.e., Word or blog) file shall be maintained to describe the events of packing, scanning, etc. See the sample log file located at: ./eng_research_TO3/Imatron/C300-Data%2006-Aug-2013/_DataSets/High_Clutter/Contents%20of%20Bag%200032/DETAILS%20OF%20BAG%200032.docx

11 Tasks before Arrival at Heartscan

1. Revise/iterate this document
2. Procure objects and tools (as listed throughout this document)
   a. Before departure for CA
   b. After arrival in CA
3. Access databases used for TO1 and TO3
4. Develop new databases for TO4

12 Initial Tasks upon Arrival at Heartscan

12.1 Object Sorting and Cataloging

1. Find work space (tables)
2. Sort out objects
3. Assure all objects labeled and labels do not duplicate
4. Catalog all objects
5. Pack and document test bag

12.2 Scanning

1. Agree on scanning and reconstruction protocols
2. Scan test bag
3. Transfer raw data
4. Perform off-line reconstruction
5. Estimate time per bag
6. Agree on total number of bags to scan during the week

12.3 Other Initial tasks

1. Agree on roles and responsibilities
2. Agree on schedule
3. Discuss need for breaks and food
4. Determine saline concentration
5. Discuss file naming conventions
6. Measure physical mass, density, volume and density of (tolerances TBD):

TO4 Scan Plan, Page 14

a. Containers
b. Targets
c. Non-targets

13 Item Responsibility

What	Who	Notes
Scale	CC	< 1 g accuracy
250 ml containers	CC	McMaster
Perishable items. Fruit, vegetables, etc.	CC	Purchase upon arrival in CA
Stream of commerce containers, > 125 ml for saline	AW	Soda cans, bottles plastic bins, metal, thermos. Most 250 ml or larger
Modeling clay. Sculpey. www.sculpey.com/products/clays/original-sculpey	CC	
Rubber sheets	CC	
Salt (w/o iodine)	AW	For doping water 1050-1150 MHU
Tape for IDs	AW	
Marker (permanent)	AW	
Laptop for documentation	AW	
Digital camera	AW	
Video camera	AW	
Plastic bags (sandwich, quart, gallon)	CC	For containing saline and storing modeling clay
Plastic wrap	AW	For wrapping modeling clay and
Knife	AW	For cutting modeling clay
Rubber hammer	AW	For shaping modeling clay
Ruler (metric, >30 cm)	AW	
Tape measure (metric, > 1 m)	AW	
Measuring cup	AW	For saline
Plastic storage bins < 30" long	CC	To scan instead of suitcases
Ethernet switch and cables	CC	
Saw	CC	For cutting rubber and pipes
Powder	HM	
Clothing for packing	AW	
Displacement gradients	RH	Used in TO1
Old laptop computer	RM	
Picture frames	RM	
Masking and duct tape	AW	
Ceramic mugs	AW	
Pipes (metal, plastic)	CC	
Rubber bands	AW	
Oil	CC	
Honey	CC	
Peanut butter	CC	

TO4 Scan Plan, Page 15

Abbreviations:

1. CC: Carl Crawford
2. AW: Alyssa White
3. HM: Harry Martz
4. RH: Rick Moore

14 Local Shopping

1. Costco/Safeway
    a. Peanut butter
    b. Coke
    c. Oil
    d. Fruit
    e. Honey
    f. Salt
    g. Pepper
    h. Rice
    i. Batteries
    j. Sugar (power)
    k. flour
2. Lowes
    a. Bins
    b. Masking tape
    c. Pipes, metal and pvc
    d. Saws
    e. Powder
    f. Misc. tools for scanning
    g. Large bottles for mixing saline
3. Goodwill
    a. Clothes
    b. Shoes/boots
    c. Concealment items
    d. Clutter items

15 Local Resources

Home Depot
2 Colma Blvd
Colma, CA
(650) 755-9600

Target

TO4 Scan Plan, Page 16

1150 El Camino Real
San Bruno, CA

Walmart
600 Showers Dr
Mountain View, CA 94040
Phone: (650) 917-0795

Lowes (closest to Heatscan)
720 Dubuque Ave
South San Francisco
(650) 452-1040

Costco
451 S Airport Blvd
S San Francisco, CA

Safeway
30 Chestnut Ave.
South San Francisco, CA

Goodwill
225 Kenwood Way
S San Francisco, CA
(650) 737-9827
http://sfgoodwill.org/

Trader Joes
765 Broadway
Millbrae, CA
(650) 259-9142

16 Additional Notes

1. This plan was written, in part, based on the lessons learned during collecting data for the Segmentation Initiative (Task Order 1) and the Reconstruction Initiative (Task Order 3). See the final reports for those projects for additional information.
2. Spares should be brought for critical equipment such as scales, cameras, digital cameras, archive material and computers.

17 Revision History

Version	Date	Author	Revisions
1	9/12/2013	Crawford	Initial release.

TO4 Scan Plan, Page 17

2	9/18/2013	Crawford	Numerous changes mainly based on phone conference on 9/130/2013.
3	9/23/2013	Crawford	Miscellaneous changes
4	9/23/2013	Crawford	Changes due to feedback from vendor, LLNL and Franco
5	9/25/2013	Crawford	Misc. last-minute changes
6	10/8/2013	Crawford	Revised after scanning took place
7	10/6/2014	Crawford	Removed revision history. Added correspondence to Imatron images.

11.3.3 Lessons Learned

"Lessons Learned from Collecting Data for the Reconstruction and ATR Projects"

Lessons Learned from Collecting Data for the Reconstruction and ATR Projects

Contents

TO4 Scanning Lessons Learned, Page 2

1 Synopsis

The purpose of this document is to document the lessons learned when collecting data for the reconstruction project (Task Order 3) and the ATR project (Task Order 4). The data were collected on the Imatron medical CT scanner located near the San Francisco airport at the Heartscan medical clinic. The points noted herein should be addressed when data is collected in the future.

2 Lessons Learned

2.1 Specifying what has to be done during data collection

1. A lexicon should be created for all aspects of data collection including target preparation, packing, scanning, data archiving and retrospective reconstruction.
2. A glossary of terms and acronyms should be made available to all personnel.
3. A printer should be made available to print out instructions.
4. Standard operating procedures (SOP) should be developed for making specimens (e.g., saline), containing targets, packing bags, scanning, data archiving and retrospective reconstruction.

2.2 Data recording

1. Using a slice serial number (SSN) is a good idea for each scan. This leads to common and sequential naming convention for raw data, corrected data and images.
2. Templates for all spreadsheets should be created in advance.
3. The fields and their contents should be specified in writing.
4. The blog was a good idea; however, additional personnel are required to record sufficient detail.
5. Consider taping to the wall each day a list of objects to be scanned along with goals for packing and in particular clutter.
6. Videos should be made for all bags with the exception of bags that are rescanned without changing contents or their locations.

2.3 Objects

1. Avoid object philosophy issues by not doing the following.
    a. Scan targets of the same type that are touching to prevent merged objects.
    b. Scan bags with objects of the same type. For example, make sure that rubber is not present in other forms (e.g., a rubber mallet).
2. Do not mix targets and pseudo targets. If pseudo targets need to be scanned, do not mix those scans with the scans of targets.

2.4 Object spreadsheet

3. All objects (targets and non-targets) should be entered in advance of scanning. The only exceptions are materials that are made on site (e.g., saline and concealed/contained objects).
4. Separate sheets should be used for source materials (e.g., big blocks of clay, rubber sheets or containers of saline), specimens, and non-targets.
5. Agree on units (e.g., mm or cm) for all fields.

TO4 Scanning Lessons Learned, Page 3

2.5 Labels

1. Different colors should be used for different types of objects (target, non-target, pseudo target, and specimen).
2. A computerized label maker should be used instead of handwriting labels.
3. Labels with numbers should be printed in advance.
4. Consider using bar codes to label objects and a bar code reader to verify that the proper objects are packed in a bag. Software may have to be written to transfer the output of the bar code reader to spreadsheets. The bar codes should be a supplement to human-readable labels.
5. Mark the preferred axis of an object on the object.
6. The SSN should be shown in all pictures of packed bins.

2.6 Imatron scanner itself

1. Figure out a way to eliminate on-line reconstruction, which limited the bag scanning rate to half of the predicted rate of approximately ten bags/hour.
2. A better metric than bags/hour is number of helical scans protocols/hour.
3. Determine how to connect a Windows 7 computer to the reconstruction computer so that projection data and images can be transferred from the scanner without the use of on intermediate computer.

2.7 Scanning on the Imatron scanner

1. Investigate other scanning protocols that would allow scanning longer bags during one helical scan operation. However, changing protocols may render previously collected data useless.
2. Drop the reconstruction dependencies. The Imatron recon is a useful QC tool, but delays each scan and requires large data space. This limits the bag scanning rate and raises the changes of data transfer/management errors
3. Scanning portions of a bag/bin can be done when it is known that nothing relevant is contained in the end a bag/bin.
4. Keep the bags aligned with the table top (i.e., do not flip, twist and rotate) so that the packing locations represent where objects are in the images.

2.8 Packing bags

1. Clear bins are useful in order to see the contents in the bin.
2. Simulate real bags with bins by adding two aluminum bars and a couple of wheels to the interior of the bins.
3. Consider the cost/benefit of real bags. Typical luggage was too long and sometimes too wide to be scanned on the Imatron CT scanner.
4. The contents and locations of objects should be specified in advance.
5. A person should check that the correct objects are placed in the bag, in the proper location and in the proper orientation. This means at least two people are responsible for packing versus the one person that was used.
6. The number of targets per bag should be an average, not absolute per bag.

TO4 Scanning Lessons Learned, Page 4

2.9 Targets

1. Consider the cost/benefit of pseudo targets; their use increases the complexity logging the scanning activity and complicates the scoring of the results of an ATR,.
2. Make sheets and one bulk the same material to eliminate the question of when does a sheet become a bulk.

2.10 Personnel supporting the scanning

1. More people would be helpful, especially for verification of samples and data recording. The verification should include, but not limited to, the following tasks, when packing and scanning a bag.
    a. All objects are properly labeled.
    b. All objects are entered properly in the database.
    c. The locations and orientations of all objects are recorded in the database.
    d. Object philosophy issues are avoided.
    e. Pictures are available for all objects.
    f. Unpacking videos are recorded.
    g. Scan information (e.g., RCP numbers) are recorded.
2. The tool developer and at least one ATR developer should be present to witness scanning and to make sure that sufficient information about the scans is recorded.
3. Two training sessions should be conducted before most of the bags are scanned one without the scanner (i.e., a virtual dry run) and one with the scanner, but only scan approximately ten bags (i.e., a real dry run).
4. Agree on roles and responsibilities of all personnel.
5. Switch roles during the scanning to understand the needs of all personnel.

2.11 Scanning site

1. Arrive a day in advance to:
    a. Prepare all liquids
    b. Sort out and label all objects.
    c. Perform a dry run with the scanner.
    d. Prearrange data extraction to demonstrate data can be extracted from the scanner, and rectify extraction issues ahead of the actual visit.
    e. Arrange for work surfaces (tables).
2. When departing:
    a. Make arrangements to ship all scanning items back to ALERT.
    b. Restore space to state it was in before arrival.

2.12 Tools used as part of data collection.

1. To the degree possible, preprint labels. This avoids duplication, and permits accounting for pre-determined blocks of numbers
2. Bring multiples of the following items so that their use does not get into the critical path cameras, pens, scales and labels.

TO4 Scanning Lessons Learned, Page 5

3. Bring a portable printer so that updated SOPs can be provided to personnel.
4. Bring a portable scanner or iPhone camera so that related documents can be archived before they are lost.

2.13 Off-line reconstruction of the data

1. Make sure that sufficient people are available to perform this task.
2. Tasks also include
3. compensating for overlap between helical protocols
    a. Renaming files
    b. Writing scripts to drive xrec
    c. Finding bounding boxes of targets. The bounding boxes should be created onsite so that the locations of the objects are fresh in the minds of the people who packed the bags.
    d. Verifying the contents of databases (spreadsheets)

2.14 Ground-truth

1. Be wary of Mevislab; it is buggy.
2. Check the literature for options for segmenting images.
3. Generate ground truth on site.
4. The rules for how to generate ground truth should be supplied in writing in advance creating the ground truth.

2.15 FTP site

1. Its organization (layout) should be specified before scanning begins.
2. Try to upload data during the scanning session.
3. Label images should be compressed with gzip.
4. If FITS format is used, introduce a suffix for compressed FITS files (e.g., .fgz).

2.16 Miscellaneous

1. Review best practices in other related fields. For example, review practices used for DNDO data collection.
2. Figure out a way to avoid the object philosophy issue. In TO3, N rubber sheets were stacked leading to a question of whether the stack was one object or N objects? In TO4, is a rubber mallet a target?
3. All data (projection, images and spreadsheets) should be consumed (used) earlier in the project. If possible, sample tools (ATR, reconstruction) should be available at the time that data are acquired.

11.4 Support Tools

11.4.1 Tools Spec

"ALERT ATR Project: Software Tools Specifications"

ALERT ATR Project: Software Tools Specifications

Author: Franco Rupcich
Email: franco.rupcich@gmail.com
Phone: 414-559-3338

Version 7

Contents

1 Introduction

This document is a technical specification and reference guide for the software tools and associated log files for the ALERT ATR project (also denoted Task Order 4 [TO4]). The information in this document supersedes that in the Top-Level Technical Specifications document. The tools that are described in this document are summarized in the following table.

Table 1: TO4 Software Tools

Executable/Program Name	Software Tool	Description
satr	Sample ATR	Sample ATR. Performs connected component labeling (CCL) on an input CT image and produces a label image. Demonstrates how to read CT images and produce label and log files.
dder	Detection Determination (scoring)	Scores an input label image produced by an ATR against the corresponding ground truth (GT) image based on specified values of precision and recall and outputs the results to two log files
pdpfa	PD/PFA Determination	Compiles the results from a specified set of images that have been scored using *dder*, determines probability of detection and probability of false alarm statistics, and writes results to four log files.
gen_pdpfa.sh	Generate PD/PFA	Runs a specified ATR on a specified set of CT images to produce a set of corresponding label images, scores the label images against the GT label images using *dder*, and then generates probability of detection (PD) and probability of false alarm (PFA) statistics using *pdpfa*.
gtver	GT Verification	Verifies GT label images by cross-checking the mass and bounding box of each target calculated from the GT label images against the mass and bounding box recorded in the object and packing databases.
mi2fits	MI to FITS File Converter	Converts an mi-formatted file to a FITS-formatted file.
fits2mi	FITS to MI File Converter	Converts a FITS-formatted file to an mi-formatted file.
raw2fits	Raw to FITS File Converter	Converts a raw-formatted file (unsigned short, short, unsigned int, int, or float) to a FITS-formatted file.
mmi	Merge CT and Label Images	Overlays a label image onto a CT image. The resultant merged mi formatted image can be viewed using xpic.

2 Definitions

See ALERT ATR Project: Top-Level Specifications for the list of terms and acronyms for this project, as well as for details regarding targets vs. pseudo-targets, detection, false alarms, misses, probability of detection, probability of false alarm, recall, and precision.

3 General

3.1 Supported Operating Systems

The tools are supported on the following operating systems:

1. 32-bit Linux
2. 64-bit Linux
3. 64-bit Mac OS X (version 10.x)

3.2 Required Packages

Below is a list of packages required to be installed on Linux and OSX before running the tools:

1. dos2unix
2. gcc
3. make
4. rcs

3.3 CRC Parse Library

The TO4 software tools rely on the parse library from the *crc* support package, a package of tools written by Carl Crawford. The parse library comes bundled with the TO4 tools. Additional information about the crc package can be found in the crc documentation, which can be found at http://www.csuptwo.com/crawford_software.html .

3.4 Command Line Functionality

Each of the tools accepts command line options and flags, all of which are optional. In other words, the tools can be run without any command line options, in which case a set of default command line options will be used.

Passing a "?" to any of the tools except *gen_pdpfa.sh* will give the help dialog. For *gen_pdpfa.sh*, the –h flag gives the help dialog.

3.5 TO4 Database

The TO4 database is a version-controlled Excel workbook comprising multiple worksheets containing information about each object and scan. The overall database is further broken down into three individual databases (each is a single spreadsheet in the TO4 database workbook):

1. Object database – contains information about each object
2. Packing database – contains information about each bag
3. Height database -- indicates the row height of the patient table for each bag

The software tools use CSV-formatted versions of the three database files. The CSV versions are created by saving an Excel spreadsheet as a .csv file. The CSV versions of the database files are distributed with the tools package and specified in the following sections.

3.5.1 Object Database

The object database contains the following information about each object in the following order:

1. ID
2. Description (in words); maximum of 10 words
3. Task Order (TO) -- the Task Order in which the object was procured
4. Type
   a. n = non-target
   b. x or ? = don't know enough information about object (probably from TO1 or TO3)
   c. t = target
   d. pt = pseudo-target
5. Sub-type
   a. E = enclosure (may contain target or non-target or be empty)
   b. c = clay
   c. r = rubber
   d. s = saline
6. Form
   a. sheet
   b. bulk
7. Merged – Only applicable for type "target." If "y", then this target is a merged target composed of the two targets indicated by the IDs in the "Parent" and "Container" fields. Only targets of the same type can be considered merged
8. Parent
   a. If "Merged" field is "n", then this is the ID of the material that was put into the container specified in the "container" field of this database. When a parent ID is present, this is considered to be a specimen (e.g., saline packed in a container). The ID should have an "S" on the label.
   b. If "Merged" field is "y", then this field is the ID of the first target composing the merged target.
9. Container
   a. If "Merged" field is "n", then this field is the ID of container used in the construction of a specimen.
   b. If "Merged" field is "y", then this field is the ID of the second target composing the merged target
10. Vendor and part number
11. Dimension 1 [mm]
12. Dimension 2 [mm]
13. Dimension 3 [mm]

14. Dimensions [mm] – can specify dimensions in X x Y x Z format
15. Volume (of container and contents) [cc]
16. Mass [g]

Note: A target may be scanned in multiple bags. A target retains the same ID across all scans in which it is present. If a target is reshaped or a liquid is put in a new container, the target will be given a new ID. For example, if a piece clay is reshaped, the clay will be given a new ID.

3.5.2 Packing Database

The packing database contains the following information about each object in each scan in the following order.

1. SSN
2. ID
3. Description (in words): A maximum of 20 words.
4. isChild? -- Indicates whether this target is contained in a merged target within this bag. This is to assure that only the merged target (and not the individual targets it is composed of) is accounted for during scoring
5. Location code: A three letter designation of the form xyz, where x, y and z are either A, B or C. The details of this code are specified in the scanning specification.
6. Orientation code: The code is x, y, z, for the object's preferred axis aligned with x, y, z axes of the bag, respectively. The code 'n' means the object is not aligned with any axis. The preferred axes of objects are defined in the scanning specification.
7. Level of Difficulty (LOD)
    a. Low
    b. High
8. x_min – bounding box xmin [pixels]
9. x_max – bounding box xmax [pixels]
10. y_min – bounding box ymin [pixels]
11. y_max – bounding box ymax [pixels]
12. z_min – bounding box zmin [pixels]
13. z_max – bounding box zmax [pixels]

3.5.3 Height Database

The height database contains the following information about each scan in the following order:

1. SSN
2. Height – row number of patient table

3.6 File Formats

3.6.1 Images

All CT, GT label, and ATR label images must be in FITS format (16-bit, unsigned integer). In addition, the files should also be compressed using *gzip*. The extension of these files is .fits.gz.

3.6.2 Log Files

Log files are written as either text (.txt) or tab-delimited (.xls). Tab-delimited (.xls) log files can be opened in Excel to facilitate filtering and sorting of information.

3.6.3 Database Files

The software tools use CSV formatted versions of the object, packing, and height database files. The CSV versions of the database files are distributed with the tools package. See Section 3.5 for more details.

3.7 Image File Naming Conventions

Filenames are a single letter followed by the SSN (zero padded to three digits), followed by the extension *fits.gz* (gzipped compressed FITS format). The letter code is as follows:

I – CT image
G – GT label image
A – ATR label image

For example, the CT, GT label, and ATR label images for SSN 50 are I050.fits.gz, G050.fits.gz, and A050.fits.gz, respectively.

NOTE: The SSNs, and thus the filenames, range from 004 to 193. However, due to corrupt/missing data, SSNs 27 and 160 are not used.

4 Building the Tools

The following instructions demonstrate how to build the tools on both Linux and OSX.

1. Create a working directory for the TO4 software tools. For the sake of clarity, we will refer to this directory as my_dir/.

2. Download the latest to4-tools tar file from the ftp site (**/eng_research_TO4/tools/**) and place it into my_dir/.

3. Untar the to4-tools tar file (tar xzvf *filename.tar.gz*). This will automatically create a directory called to4-tools_vXX/, where XX is the version number. In addition, a soft-link called to4-tools/ will be created. This soft-link will always point to the most recent to4-tools_vXX/ directory.

4. cd into to4-tools/

5. Type "make" to build the tools. In addition, this will automatically create four directories in my_dir/, but only if they have not already been created (i.e., these directories will not be overwritten upon subsequent builds). These four directories are the default directories used by *gen_pdpfa.sh*. They are

   a. ct/ --> *gen_pdpfa.sh* will look for CT images in this directory by default
   b. gt/ --> *gen_pdpfa.sh* will look for GT label images in this directory by default
   c. labels/ --> *gen_pdpfa.sh* will write ATR label images to this directory by default
   d. logs/ --> *gen_pdpfa.sh* will write log files to this directory by default

6. Repeat Steps 2-5 each time a new to4-tools tar file is uploaded to the FTP site.

5 Generating PD/PFA Results Using gen_pdpfa.sh

5.1 Overview

gen_pdpfa.sh is a (bash) shell script that performs the following functions:

1. Run a user-specified ATR (*satr* by default) on a user-specified set of CT images to produce a set of corresponding label images

2. Run *dder* to score the label images produced in Step 1 against the respective ground truth (GT) label images

3. Run *pdpfa* using results produced in Step 2 to generate probability of detection (PD) and probability of false alarm (PFA) statistics for the set of specified CT images. These statistics are output to several log files (see Section 7.3 for more details on the *pdpfa* log files).

Researchers are required to run this script so that all ATR results may be reported in a standardized format.

5.2 Running gen_pdpfa.sh

The following instructions describe how to run *gen_pdpfa.sh* with no command line options (i.e., using the default behavior). (For more information regarding specifying command line options for *gen_pdpfa.sh*, see Section 6.4.) For the sake of clarity, we will assume the user has a working directory named my_dir/.

1. Build the latest version of the software tools (see Section 4)

2. Download the CT images from the FTP site and place them into my_dir/ct/

3. Download the latest GT label images from the FTP site and place them into into my_dir/gt/

4. cd into my_dir/to4-tools/

5. Run ./gen_pdpfa.sh

By default, *gen_pdpfa.sh* will run *satr* for all CT images in my_dir/ct/ (for which a GT label image exists in my_dir/gt/), and it will place all label images produced by *satr* into my_dir/labels/. *dder* will then score each label image in my_dir/labels/ against the corresponding GT label image in my_dir/gt/ directory.

Log files produced by *satr* (or the specified ATR program) will be placed into my_dir/logs/atr_logs. Log files produced by *dder* will be placed into my_dir/logs/dder_logs. *pdpfa* will compile the results from the *dder* log files and generate PD and PFA log files, which will be placed into my_dir/logs/pdpfa_logs. *gen_pdpfa.sh* also generates a log file containing the command line options used, as well as the standard output of calls to *satr*, *dder*, and *pdpfa*. The *gen_pdpfa.sh* log file will be placed in my_dir/logs/pdpfa_logs. Descriptions of all log files can be found in Section 7.

5.3 Standardized Reporting

The five log files located in my_dir/logs/pdpfa_logs after running *gen_pdpfa.sh* are the ones used for standardized reporting for this project (NOTE: the following list shows the default log file names):

1. gen_pdpfa_log.txt
2. pdpfa_log_summary.txt
3. pdpfa_log_pds.xls
4. pdpfa_log_detections.xls
5. pdpfa_log_false_alarms.xls

6 Tools Specifications

Each of the software tools described in this section allows for input of command line options and flags. If no options are provided, the program will run with a set of default arguments (specified in the Options table for each software tool). The default behavior of each program can be changed using the options and flags described in the following sub-sections.

6.1 satr

6.1.1 Synopsis

satr is a "sample ATR" program. It takes as an input a CT image, performs erosion (using a 3x3x3 kernel) and connected component labeling (CCL) on it, and outputs a label image and a log file. *satr* calculates the mass of objects and has a classifier based only on mass and density.

satr was originally written by Seemeen Karimi [3] and modified by Carl Crawford for this project.

6.1.2 Input Files

1. CT image [.fits.gz, .mi]

6.1.3 Output Files

1. Label image [.fits.gz, .mi]
2. Log file [.txt]

6.1.4 Standard Output

1. Program name
2. Start time
3. RCS ID
4. File format
5. CT image filename
6. Label image filename
7. Log filename
8. Height database filename
9. Number of rows in images
10. Number of columns in images
11. Number of slices in images

12. Number of slices processed
13. FOV [mm]
14. Pixel size [mm]
15. Offset [MHU]
16. Slice Spacing [mm]
17. Minimum mass [g]
18. Lower threshold [MHU]
19. Upper threshold [MHU]
20. CCL delta [MHU]
21. Connectivity type
22. Number of labels produced
23. End time

6.1.5 Flags

Flag	Synopsis
?	Print the help dialog
-d	Increment the debug flag
-m	Use .mi file format instead of FITS
-p	Do not print the program information to the command line

6.1.6 Options

Option	Default Argument	Synopsis
if	ct.fits.gz	Filename of input CT image. By default, this file must be a gzipped FITS file (i.e., extension is .fits.gz). If no extension is specified in the argument and the –m option is not used, then .fits.gz will automatically be appended. If the –m option is used, then .mi will automatically be appended.
of	label.fits.gz	Filename of output label image. By default, the file will be written as a gzipped FITS file (i.e., extension is .fits.gz). If no extension is specified in the argument and the –m option is not used, then .fits.gz will automatically be appended. If the –m option is used, then .mi will automatically be appended.
logfn	log.txt	Filename of output log file.
hdbf	hdb.csv	Filename of height database file (.csv format). This file contains the row height (in pixels) of the patient table for each SSN, which is used to zero out below the patient table. This file comes packaged with the tools and is located in to4-tools/dbase/hdb.csv.
spacing	1.5	Spacing between slices [mm].
first	1 (first slice in image)	The slice of the image at which to begin processing.
count	0 (process all slices)	Number of slices to process
fov	475	Image field of view [mm]
offset	0	Image offset [MHU]. This value is subtracted from each voxel before running CCL. It should be used in the case of an image that does not have air = 0 MHU.
lt	1000	Lower threshold for acceptable pixel values used during labeling [MHU]

Option	Default Argument	Synopsis
ht	2000	Upper threshold for acceptable pixel values used during labeling [MHU]
ccldelta	100	Maximum absolute value of the difference between neighbor values used during merging [MHU]
minmass	50	Minimum mass to be considered a label [g]
connectivity	0	Connectivity type. For 2D: 0=edge, 1=edge and vertex. For 3D: 0=face, 1=face, edge, and vertices.

6.2 dder

6.2.1 Synopsis

dder is the "detection determination" program. It scores an input label image produced by an ATR against the corresponding GT image based on specified values of precision and recall and outputs the results to two log files.

6.2.2 Input Files

1. ATR label image [.fits.gz]
2. GT label image [.fits.gz]
3. Object database file [.csv]
4. Packing database file [.csv]

6.2.3 Output Files

1. dder summary log file [.txt] – summary scoring information
2. dder false alarm log file [.xls] – information for each false alarm

6.2.4 Standard Output

1. Program name
2. Start time
3. RCS ID
4. ATR label image filename
5. GT label image filename
6. Object database filename
7. Packing database filename
8. Summary log filename
9. False alarm log filename
10. Precision (target, bulk)
11. Recall (target, bulk)
12. Precision (target, sheet)
13. Recall (target, sheet)
14. Precision (pseudo-target, bulk)
15. Recall (pseudo-target, bulk)
16. Precision (pseudo-target, sheet)
17. Recall (pseudo-target, sheet)
18. Alpha parameter
19. End time

6.2.5 Flags

Flag	Synopsis
?	Print the help dialog
-p	Do not print the program information to the command line

6.2.6 Options

Option	Default Argument	Synopsis
ifn_atr	atr.fits.gz	Filename of input label image. This file must be a gzipped FITS file (i.e., extension is .fits.gz). If no extension is specified in the argument, then .fits.gz will automatically be used.
ifn_gt	gt.fits.gz	Filename of input GT label image. This file must be a gzipped FITS file (i.e., extension is .fits.gz). If no extension is specified in the argument, then .fits.gz will automatically be used.
ifn_odb	odb.csv	Filename of object database (.csv format). This file comes packaged with the tools and is located in to4-tools/dbase/odb.csv.
ifn_pdb	pdb.csv	Filename of packing database (.csv format). This file comes packaged with the tools and is located in to4-tools/dbase/pdb.csv.
ofn_log	dder_log	Basename used in generating the filenames for the two log files output by *dder*. The argument should not include an extension. The SSN will be parsed from the input GT label image filename and automatically appended to the basename of the two log files, along with each log file's suffix and extension. For example, if the basename is "dder_log", and *dder* is being run for SSN 6, then the two log file names will be dder_log_summary_006.txt and dder_log_false_alarms_006.xls.
p_bulk	0.5	Precision for bulk targets
r_bulk	0.5	Recall for bulk targets
p_sheet	0.2	Precision for sheet targets
r_sheet	0.2	Recall for sheet targets
p_pt_bulk	0.5	Precision for bulk pseudo-targets
r_pt_bulk	0.5	Recall for bulk pseudo-targets
p_pt_sheet	0.1	Precision for sheet pseudo-targets
r_pt_sheet	0.1	Recall for sheet pseudo-targets

Option	Default Argument	Synopsis
alpha	0.0	Multiplier used for determining when an "incomplete detection" occurs. An incomplete detection occurs when the calculated precision and recall are both greater than or equal to alpha times the specified precision and recall (where the specified precision and recall are p_bulk, r_bulk, p_sheet, r_sheet, p_pt_bulk, r_pt_bulk, p_pt_sheet, and r_pt_sheet). Incomplete detections are reported in the *dder* log files. **NOTE:** Incomplete detections still count as false alarms, since they do not meet the specified precision/recall for a detection. The ATR labels that generate an incomplete detection can be found under the "False Alarms" section of the *dder* Summary Log File, as well as in the *dder* False Alarms Log File. An incomplete detection is any false alarm reported in the log file that has intersecting GT labels reported. In other words, intersecting GT labels for a false alarm are only reported if the ATR label meets the specified precision/recall for an incomplete detection (alpha*p, alpha*r), AND the ATR label and GT label intersect by at least one pixel (for the case in which alpha = 0).

6.3 pdpfa

6.3.1 Synopsis

pdpfa compiles the results from a specified set of images that have been scored using *dder*, determines probability of detection and probability of false alarm statistics, and writes results to four log files.

6.3.2 Input Files

1. dder log list file (contains list of *dder* log files from which to read scoring information) [.txt]
2. Object database file [.csv]
3. Packing database file [.csv]

6.3.3 Output Files

1. pdpfa summary log file [.txt] – summary PD/PFA information
2. pdpfa detections log file [.xls] – indicates whether each target was detected or missed
3. pdpfa pds log file [.xls] – PD statistics
4. pdpfa false alarm log file [.xls] – information for each false alarm produced

6.3.4 Standard Output

1. Program name
2. Start time
3. RCS ID
4. Input list filename
5. Object database filename
6. Packing database filename
7. Output log basename
8. Output summary log filename
9. Output detections log filename

10. Output false alarm log filename
11. Output PD log filename
12. End time

6.3.5 Flags

Flag	Synopsis
?	Print the help dialog
-p	Do not print the program information to the command line

6.3.6 Options

Option	Default Argument	Synopsis
ifn_list	list.txt	Filename of .txt file containing the filenames of the *dder* summary log files from which to read scoring data. Each filename should be printed on a separate line, e.g.: /home/franco/to4-tools/logs/dder_log_summary_006.txt /home/franco/to4-tools/logs/dder_log_summary_007.txt /home/franco/to4-tools/logs/dder_log_summary_008.txt The probability of detection and probability of false alarm statistics calculated by *pdpfa* are for only those SSNs corresponding to the *dder* log files specified in this list file.
ifn_odb	odb.csv	Filename of object database (.csv format). This file comes packaged with the tools and is located in to4-tools/dbase/odb.csv.
ifn_pdb	pdb.csv	Filename of packing database (.csv format). This file comes packaged with the tools and is located in to4-tools/dbase/pdb.csv.
ofn_log	pdpfa_log	Basename used in generating the filenames for the four log files output by *pdpfa*. This argument should not include an extension.

6.4 gen_pdpfa.sh

6.4.1 Synopsis

gen_pdpfa.sh is a (bash) shell script that runs a specified ATR on a specified set of CT images to produce a set of corresponding label images, scores the label images against the ground truth (GT) label images using *dder*, and then generates probability of detection (PD) and probability of false alarm (PFA) statistics using *pdpfa*.

6.4.2 Input Files

1. CT image(s) [.fits.gz]
2. GT image(s) [.fits.gz]
3. Label image(s) [.fits.gz]
4. Object database file [.csv]
5. Packing database file [.csv]
6. Height Database file [.csv]

7. List of SSNs for which to calculate PD and PFA statistics (This file is optional. By default, *gen_pdpfa.sh* will run on all SSNs for which both a CT and GT label image exist in the specified directories).

6.4.3 Output Files

Because *gen_pdpfa.sh* runs *satr*, *dder*, and *pdpfa*, its output includes log files generated by those programs. The following log files are generated by *satr*, *dder*, and *pdpfa* after they are called by *gen_pdpfa.sh*:

1. satr summary log file for each SSN [.txt]
2. dder summary log file for each SSN [.txt]
3. dder false alarm log file for each SSN [.xls]
4. pdpfa summary log file [.txt]
5. pdpfa detections log file [.xls]
6. pdpfa pds log file [.xls]
7. pdpfa false alarm log file [.xls]

The following log files are generated by *gen_pdpfa.sh* itself:

8. dder log list file [.txt] – list of *dder* log files used by *pdpfa* to compile results
9. gen_pdpfa.sh log file [.txt] – standard output from *satr*, *dder*, and *pdpfa*

6.4.4 Standard Output

1. Program name
2. Start time
3. CT directory
4. GT directory
5. Label directory
6. Logs directory
7. Label image prefix
8. SSN list filename
9. ATR executable filename
10. ATR args
11. Precision (target, bulk)
12. Recall (target, bulk)
13. Precision (target, sheet)
14. Recall (target, sheet)
15. Precision (pseudo-target, bulk)
16. Recall (pseudo-target, bulk)
17. Precision (pseudo-target, sheet)
18. Recall (pseudo-target, sheet)
19. Alpha
20. End time

6.4.5 Flags

Flag	Synopsis
-h	Print the help dialog

6.4.6 Options

Option	Default Argument	Synopsis
-c	../ct/ (this directory is automatically created by the top-level "Makefile," and resides one level above the to4-tools/ directory).	Directory containing CT images
-g	../gt/ (this directory is automatically created by the top-level "Makefile," and resides one level above the to4-tools/ directory).	Directory containing GT label images
-l	../labels/ (this directory is automatically created by the top-level "Makefile," and resides one level above the to4-tools/ directory).	Directory to which ATR label images are written
-o	../logs/ (this directory is automatically created by the top-level "Makefile," and resides one level above the to4-tools/ directory).	Directory to which *satr*, *dder*, *pdpfa*, and *gen_pdpfa.sh* log files will be written
-x	A	Prefix to be used for each label image produced by *satr*. When *satr* is run, it will use this prefix along with the SSN to create the full label image filename. For example, if the prefix "A" is used, then the label image filename for SSN 6 will be A006.fits.gz.
-a	./satr/satr	Relative or absolute pathname of ATR executable
-z	None (no additional arguments are passed to the atr)	List of additional options to pass to the specified ATR. These options and their arguments must be enclosed inside double quotes (" "). For example: ./gen_pdpfa.sh –z "lt=1500 ht=2200 ccldelta=75" **NOTE:** satr/atr options for specifying the input CT image (if), output label image (of), output log file (logf), and height database filename (hdbf) should NOT be included in the argument for the -z option. This is because the first three options' arguments change as satr/atr is called on each SSN. In addition, the height database file is passed as a separate argument rather than being included as apart of the –z option.

Option	Default Argument	Synopsis
-f	None (program will run for all CT and GT label images that exist in the specified directories)	Filename of .txt file containing list of SSNs for which to calculate PD and PFA statistics. Each SSN should be printed on a separate line, e.g.: 6 7 8 23 188
-p	0.5	Precision for bulk targets
-r	0.5	Recall for bulk targets
-q	0.2	Precision for sheet targets
-s	0.2	Recall for sheet targets
-t	0.5	Precision for bulk pseudo-targets
-v	0.5	Recall for bulk pseudo-targets
-u	0.1	Precision for sheet pseudo-targets
-w	0.1	Recall for sheet pseudo-targets
-b	0.0	Alpha parameter (for incomplete detections). See NOTE for this argument in the table in Section 6.2.6.

6.4.7 Notes

While the default ATR is *satr*, you may instead specify your own ATR executable using the –a option. The –z option can then be used to specify additional options and arguments to pass to your ATR. However, there is a caveat to using the –z option. Because the filenames of the input CT image, output ATR label image, and output log file change with each call to *satr* within *gen_pdpfa.sh* (i.e., for each SSN), these arguments are updated within a loop, and thus cannot be used as part of the –z option. In addition, the *satr* option "hdbf," which specifies the height database file, also should not be used as part of the –z option. **In effect, this means that your ATR must use the same command line options as *satr* for the following four options: if, of, logf, and hdbf.** See satr.c for how the code is implemented. *gen_pdpfa.sh* may later be updated to eliminate this restriction of the –z and –a options.

By default, *gen_pdpfa.sh* will run on all SSNs for which both a CT and GT image exist in the specified ct and gt directories. You may also specify a subset of SSNs to run by using the –f option to specify a file containing only the SSNs you want to run. You can use the "SSN Filter" tab of the TO4 database spreadsheet to help determine subsets of SSNs that may be useful in testing your ATR.

6.5 gtver

6.5.1 Synopsis

Verifies GT label images by checking the following criteria for a given SSN:

1. CT image and GT label image are same size (number of rows, columns, and slices)
2. GT label IDs match target IDs from packing database
3. Target masses calculated from GT label image and CT image are within +/-50% of those reported in the object database
4. GT label bounding boxes and target bounding boxes in packing database have precision and recall of at least 0.5.

6.5.2 Input Files

1. CT image
2. GT label image
3. Object database file
4. Packing database file

6.5.3 Output Files

1. gtver log file

6.5.4 Standard Output

1. Program name
2. Start time
3. CT filename
4. GT filename
5. Object database filename
6. Packing Database filename
7. Number of errors
8. Verification summary (PASSED or FAILED)
9. End time

6.5.5 Flags

Flag	Synopsis
?	Print the help dialog
-p	Do not print the program information to the command line

6.5.6 Options

Option	Default Argument	Synopsis
ifn_ct	ct.fits.gz	Filename of input CT image. This file must be a gzipped FITS file (i.e., extension is .fits.gz). If no extension is specified in the argument, then .fits.gz will automatically be used.
ifn_gt	gt.fits.gz	Filename of input GT label image. This file must be a gzipped FITS file (i.e., extension is .fits.gz). If no extension is specified in the argument, then .fits.gz will automatically be used.
ifn_odb	odb.csv	Filename of object database (.csv format). This file comes packaged with the tools and is located in to4-tools/dbase/odb.csv.
ifn_pdb	pdb.csv	Filename of packing database (.csv format). This file comes packaged with the tools and is located in to4-tools/dbase/pdb.csv.
bn_log	gtver_log	Basename used in generating the log filename. The argument should not include an extension. The SSN will be parsed from the input GT label image filename and automatically appended to the basename of the log file. For example, if the basename is "gtver_log", and *gtver* is being run for SSN 6, then the log file nameswill be gtver_log_006.txt.

6.6 mi2fits

6.6.1 Synopsis

Converts an mi-formatted image to a compressed (gzipped) FITS formatted image.

6.6.2 Input Files

1. mi-formatted image

6.6.3 Output Files

1. Compressed (gzipped) FITS formatted image

6.6.4 Standard Output

1. Program name
2. Start time
3. RCS ID
4. Input image filename
5. Input image [#rows #cols #slices]
6. First slice read from image
7. Last slice read from image
8. Number of slices converted
9. Output image filename
10. Output image [#rows #cols #slices]
11. End time

6.6.5 Flags

Flag	Synopsis
?	Print the help dialog
-p	Do not print the program information to the command line

6.6.6 Options

Option	Default Argument	Synopsis
ifn	image.mi	Filename of input mi image
ofn	*[derived from ifn arg]*.fits.gz	Filename of output gzipped FITS formatted image. Default output filename is same as input filename, but with fits.gz extension.
pix	0.928	Image pixel size [mm]
first	1 (first slice in image)	The slice of the image at which to begin processing
count	0 (process all slices)	Number of slices to convert
bitpix	1	FITS file pixel data type (1=16-bit ushort, 2=32-bit float)

6.7 fits2mi

6.7.1 Synopsis

Converts a FITS formatted image to an mi formatted image.

6.7.2 Input Files

1. FITS formatted image (.fits or .fits.gz)

6.7.3 Output Files

1. mi-formatted image

6.7.4 Standard Output

1. Program name

2. Start time
3. RCS ID
4. Input image filename
5. Input image [#rows #cols #slices]
6. First slice read from image
7. Last slice read from image
8. Number of slices converted
9. Output image filename
10. Output image [#rows #cols #slices]
11. End time

6.7.5 Flags

Flag	Synopsis
?	Print the help dialog
-p	Do not print the program information to the command line

6.7.6 Options

Option	Default Argument	Synopsis
ifn	image.fits	Filename of input FITS image
ofn	*[derived from ifn arg]*.mi	Filename of output mi image. Default output filename is same as input filename, but with mi extension.
first	1 (first slice in image)	The slice of the image at which to begin processing
count	0 (process all slices)	Number of slices to convert

6.8 raw2fits

6.8.1 Synopsis

Converts a raw-formatted (a.k.a. "flat-file") to a compressed (gzipped) FITS formatted file.

6.8.2 Input Files

1. Raw-formatted image (unsigned short, short, unsigned int, int, or float)

6.8.3 Output Files

1. Compressed (gzipped) FITS formatted image

6.8.4 Standard Output

1. Program name
2. Start time
3. RCS ID
4. Input image filename
5. Input image [#rows #cols #slices]
6. Input image raw datatype
7. First slice read from image
8. Last slice read from image
9. Number of slices converted
10. Output image filename
11. Output image [#rows #cols #slices]
12. End time

6.8.5 Flags

Flag	Synopsis
?	Print the help dialog
-p	Do not print the program information to the command line

6.8.6 Options

Option	Default Argument	Synopsis
ifn	image.raw	Filename of raw formatted image
ofn	*[derived from ifn arg]*.fits.gz	Filename of output FITS formatted image. Default output filename is same as input filename, but with .fits.gz extension.
dtype	0	Raw data type (0=ushort, 1=short, 2=uint, 3=int, 4=float)
nr	512	Number of rows in input image
nc	512	Number of columns in input image
ns	1	Number of slices to convert
first	1 (first slice in image)	The slice of the image at which to begin processing
count	0 (process all slices)	Number of slices to convert
bitpix	1	FITS file pixel data type (1=16-bit ushort, 2=32-bit float)

6.9 mmi

6.9.1 Synopsis

Overlays a label image onto a CT image. The resultant merged mi formatted image can be viewed using xpic with the "wl –o" option.

6.9.2 Input Files

1. CT image (mi format)
2. Label image (mi format)

6.9.3 Output Files

1. Merged image (mi format)

6.9.4 Standard Output

1. Program name
2. Start time
3. RCS ID
4. CT image filename
5. Number of slices used in CT image
6. Label image filename
7. Number of slices used in label image
8. First slice used in label image
9. Label image offset
10. Merged image filename
11. Number of slices processed
12. Mode used (whole, contour, or edge)
13. Multi-color (yes or no)
14. Number of labels

15. End time

6.9.5 Flags

Flag	Synopsis
?	Print the help dialog
-c	Enables contours mode
-e	Enables edge mode (default)
-m	Disables multi-color
-w	Uses whole labels
-p	Do not print the program information to the command line

6.9.6 Options

Option	Default Argument	Synopsis
cf	ct.mi	Filename of mi formatted CT image
lf	label.mi	Filename of mi formatted label image
of	merge.mi	Filename of mi formatted output merged image
offset	0	Image offset [MHU]. This value is subtracted from each voxel before processing. It should be used in the case of an image that does not have air = 0 MHU.
ctfirst	1 (first slice in image)	The slice of the ct image at which to begin processing
labelfirst	1 (first slice in image)	The slice of the label image at which to begin processing
first	1 (first slice in image)	The slice of the ct image at which to begin processing
count	0 (process all slices)	Number of slices to convert

7 Log Files

gen_pdpfa.sh, *satr*, *dder*, and *pdpfa* each produce their own log file(s), the contents of which are explained below. The names used below are the default names for the log files used by each program. The "XXX" represents the SSN number corresponding to that log file.

7.1 satr

7.1.1 satr_log_XXX.txt

Contains information about the CT image, atr, and labels produced for this SSN. See Appendix A: Example satr Log File for an example log file.

7.2 dder

7.2.1 dder_log_summary_XXX.txt

Contains score information for this SSN, including information about all targets, labels, detections, incomplete detections, false alarms, and misses. See Appendix B: Example dder Summary Log File for an example log file.

7.2.2 dder_log_false_alarms_XXX.xls

Tab-delimited log file that can be opened in Excel for filtering/sorting. Contains information about each false alarm produced by the ATR for this SSN, including the ATR label IDs and any targets or pseudo-targets that intersected each label, as well as information about any intersecting targets or pseudo-targets. The following table describes each field. See Appendix C: Example dder False Alarms Log File for an example log file.

Field Name	Description
SSN	SSN of the scored image
FA_Label_ID	Label ID of false alarm
Intersecting_Target_ID	ID of intersecting target (if applicable)
Intersecting_Target_Form	Form (bulk or sheet) of intersecting target (if applicable)
Intersecting_Target_Subtype	Subtype (saline, clay, rubber, or powder) of intersecting target (if applicable)
Precision	Precision of intersecting target (if applicable)
Recall	Recall of intersecting target (if applicable)

7.3 pdpfa

7.3.1 pdpfa_log_summary.txt

Contains summary score information for all specified SSNs, including:

1. PFA
2. Average number of false alarms per bags with at least one false alarm
3. PD based on type (target, pseudo-target, or both)
4. PD based on level of difficulty (low and high)
5. PD based on sub-type (clay, rubber, saline, and powder)
6. PD based on form (bulk and sheet)

See Appendix D: Example pdpfa Summary Log File for an example log file.

7.3.2 pdpfa_log_detections.xls

Tab-delimited log file that can be opened in Excel for filtering/sorting. Contains information for each target and pseudo-target that was scored, including whether it was detected or missed. The following table describes each field. See Appendix E: Example pdpfa Detections Log File for an example log file.

Field Name	Description
SSN	SSN of the image
Target ID	Target ID
Detected	0 or 1, indicating that the target was missed or detected, respectively
Material Type	t=target, pt=pseudo-target
Material Subtype	saline, clay, rubber, or powder
Material Form	bulk or sheet
Difficulty	low or high
Mass [g]	Mass of target [g]
Volume [cc]	Volume of target (including container and contents) [cc]

Field Name	Description
Dim x [mm]	x dimension of target [mm]
Dim y [mm]	y dimension of target [mm]
Dim z [mm]	z dimension of target [mm]
Bbox-x-min	Minimum x-coordinate of target's bounding box
Bbox-x-max	Maximum x-coordinate of target's bounding box
Bbox-y-min	Minimum y-coordinate of target's bounding box
Bbox-y-max	Maximum y-coordinate of target's bounding box
Bbox-z-min	Minimum z-coordinate of target's bounding box
Bbox-z-max	Maximum z-coordinate of target's bounding box
Object db description	Description of target taken from object database
Packing db description	Description of target taken from packing database

7.3.3 pdpfa_log_pds.xls

Tab-delimited log file that can be opened in Excel for filtering/sorting. Contains grouped PD information based on level of difficulty, form, and sub-type. Also contains PFA and average number of false alarms per bag with at least one false alarm. This log file contains information about targets only (pseudo-targets are excluded). The following table describes each field. See Appendix F: Example pdpfa PD Log File for an example log file.

Field Name	Description
Target Subtype or Form	Saline, Clay, Rubber, Bulk, Sheet, or All
Level of Difficulty	Low, High, or All
Num Targets	Number of targets in the specified grouping
Num Detected	Number of targets detected in the specified grouping
PD (targets only) [%]	PD for the specified grouping
Num Non-targets	Total number of non-targets
Num FAs	Total number of false alarms generated
PFA [%]	PFA
Num Scans with FAs	Number of scans that generated at least one false alarm
Avg Num FAs	Average number of false alarms per bags with at least one false alarm (calculated as [Num FAs] / [Num Scans with FAs])

7.3.4 pdpfa_log_false_alarms.xls

Tab-delimited log file that can be opened in Excel for filtering/sorting. Contains information about each false alarm produced by the ATR for all specified SSNs, including the ATR label IDs and any targets or pseudo-targets that may have intersected each label, as well as information about any intersecting targets or pseudo-targets. The following table describes each field. See Appendix G: Example pdpfa False Alarms Log File.

Field Name	Description
SSN	SSN of the scored image
FA Label ID	Label ID of false alarm
Intersecting Target ID	ID of intersecting target (if applicable)
Intersecting Target Form	Form (bulk or sheet) of intersecting target (if applicable)
Intersecting Target Subtype	Subtype (saline, clay, rubber, or powder) of intersecting target (if applicable)
Precision	Precision of intersecting target (if applicable)
Recall	Recall of intersecting target (if applicable)

7.4 gen_pdpfa.sh

7.4.1 gen_pdpfa_log.txt

Contains information about *gen_pdpfa.sh*, including command line options used and standard output from each call to the specified ATR, *dder*, and *pdpfa*. See Appendix H: Example of gen_pdpfa.sh Summary Log File for an example log file.

7.5 gtver

7.5.1 gtver_log_XXX.txt

Contains information about *gtver*, including command line options used, standard output, and verification results. See Appendix I: Example of gtver Log FileAppendix H: Example of gen_pdpfa.sh Summary Log File for an example log file

8 References

[1] Rupcich, F. and Crawford, C. R., "ALERT ATR Project: Top-Level Technical Specifications," Version 4, April 12, 2014.

[2] Crawford, C. R., "ATR Project Level of Difficulty Specification," February 5, 2014.

[3] Karimi, Seemeen. "Sample Segmentation Software for Segmentation Grand Challenge," April 30, 2010.

9 Revision History

Version	Changes
1	Initial revision
2	Revised based on updates to gen_pdpfa.sh and feedback from Carl Crawford
3	Added more detail
4	Revised sections and added synopses
5	Added more detail and additional tools. Revised based on feedback from Carl. Added references section. Removed terms and point to Top-Level Spec instead.
6	Added updates to gen_pdpfa.sh. Filled in section for gtver and added gtver log file.
7	Indicated that false alarms can also be found in dder false alarm log file

Appendix A: Example satr Log File

[Performer] Carl Crawford, Csuptwo
[Date]: Mon Mar 24 21:21:25 2014
[Time]: Mon Mar 24 21:21:25 2014
[CT-name] /home/franco/to4/ct/I100.fits.gz
[CT-format] FITS
[CT-columns] 512
[CT-rows] 512
[CT-slices] 244
[CT-first] 1
[CT-count] 244
[CT-fov] (mm) 475.00
[CT-pixel] (mm) 0.93
[CT-slice-space] (mm) 1.50
[CT-offset] (MHU) 0
[CT-dimension-z] (mm) 366.00
[CT-mean] (MHU) 4.36
[CT-mass] (g) 351.96
[Label-name] /home/franco/to4/labels/A100.fits.gz
[Label-format] FITS (16-bit unsigned short)
[OS] Linux
[Executable] satr
[Version] $Id: satr.c,v 1.4 2014/02/08 17:23:59 franco Exp franco $
Total-labels includes label (0) for background

[Total-labels] 2

**** satr program variables ****
#min mass (g) = 50.00
#low threshold (MHU) = 1000
#high threshold (MHU) = 2000
#ccl delta (MHU) = 100
#connectivity = 0

Label-num=0 is the background
[Label-num] 0
[Label-id] 0
[Slice-first] 1
[Slice-last] 244
[Row-first] 1
[Row-last] 512
[Column-first] 1
[Column-last] 512
[Dimension-x] (mm) 475.00
[Dimension-y] (mm) 475.00
[Dimension-z] (mm) 366.00

[Voxels] 63832324
[Mass] (g) 168.14
[Volume] (cc) 82409.87
[Mean] (MHU) 2.09
[Standard-deviation] (MHU) 56.34

[Label-num] 1
[Label-id] 1
[Slice-first] 48
[Slice-last] 138
[Row-first] 178
[Row-last] 220
[Column-first] 132
[Column-last] 327
[Dimension-x] (mm) 181.84
[Dimension-y] (mm) 39.89
[Dimension-z] (mm) 136.50
[Voxels] 130812
[Mass] (g) 183.82
[Volume] (cc) 168.88
[Mean] (MHU) 1114.57
[Standard-deviation] (MHU) 59.36

Appendix B: Example dder Summary Log File

NOTE: Fields for which there is no information recorded in the object database are marked "None Reported"

[Program-name] dder
[Version] 1.18
[Date] 05/10/14
[Time] 11:59:19
[Summary-log-file-name] /home/franco/to4/logs/v35/dder_logs/dder_log_summary_008.txt
[False-alarm-log-file-name] /home/franco/to4/logs/v35/dder_logs/dder_log_false_alarms_008.xls

Summary

[Scan-serial-number] 008
[GT-label-image-filename] /home/franco/to4/gt/G008.fits.gz
[ATR-label-image-filename] /home/franco/to4/labels/A008.fits.gz
[Num-targets] 2
[Target-ids-present] 6004 6002
[Num-target-detections] 1
[Target-ids-detected] 6002
[Num-target-misses] 1
[Target-ids-missed] 6004
[Num-pseudo-targets] 1
[Pseudo-target-ids-present] 6026
[Num-pseudo-target-detections] 0
[Num-pseudo-target-misses] 1
[Pseudo-target-ids-missed] 6026
[Num-atr-labels] 6
[ATR-label-ids] 1 2 3 4 5 6
[Num-false-alarms] 5
[False-alarm-label-ids] 6 5 4 3 1

Command Line Information

[Target-database-filename] /home/franco/to4/to4-tools/dbase/odb.csv
[Bag-database-filename] /home/franco/to4/to4-tools/dbase/pdb.csv
[GT-label-image-filename] /home/franco/to4/gt/G008.fits.gz
[GT-label-image-size (nrow ncol nslice)] [512 512 482]
[ATR-label-image-filename] /home/franco/to4/labels/A008.fits.gz
[ATR-label-image-size (nrow ncol nslice)] [512 512 482]
[Precision-bulk-target] 0.50
[Recall-bulk-target] 0.50
[Precision-sheet-target] 0.20
[Recall-sheet-target] 0.20
[Precision-bulk-pseduo-target] 0.50
[Recall-bulk-pseudo-target] 0.50

[Precision-sheet-pseudo-target] 0.20
[Recall-sheet-pseduo-target] 0.20
[Alpha] 0.00

Information for GT label image (targets and pseduo-targets)

[Number-of-targets] 2

[Target-num] 1
[Target-id] 6002
[Label-pixels] 199222
[Label-volume (cc)] 257.4
[Label-column-first] 249
[Label-column-last] 409
[Label-row-first] 257
[Label-row-last] 314
[Label-slice-first] 102
[Label-slice-last] 146
[Label-dimension-row (mm)] 52.9
[Label-dimension-col (mm)] 148.5
[Label-dimension-slice (mm)] 66.0
[Target-object-database-description] Breast Milk Bottle 5% Saline
[Target-material-form] bulk
[Target-material-subtype] saline
[Target-dimension-x (mm)] None Recorded
[Target-dimension-y (mm)] None Recorded
[Target-dimension-z (mm)] None Recorded
[Target-mass (g)] 253.0
[Target-volume (cc)] None Recorded
[Target-packing-database-description] saline
[Target-xmin] 251
[Target-xmax] 408
[Target-ymin] 255
[Target-ymax] 314
[Target-zmin] 102
[Target-zmax] 146
[Target-level-of-difficulty] low
[Target-location-code] cbb
[Target-orientation-code]

[Target-num] 2
[Target-id] 6004
[Label-pixels] 382579
[Label-volume (cc)] 494.2
[Label-column-first] 361
[Label-column-last] 443
[Label-row-first] 230
[Label-row-last] 375

[Label-slice-first] 137
[Label-slice-last] 190
[Label-dimension-row (mm)] 134.6
[Label-dimension-col (mm)] 76.1
[Label-dimension-slice (mm)] 79.5
[Target-object-database-description] Rubber Mallet
[Target-material-form] bulk
[Target-material-subtype] rubber
[Target-dimension-x (mm)] None Recorded
[Target-dimension-y (mm)] None Recorded
[Target-dimension-z (mm)] None Recorded
[Target-mass (g)] 1025.0
[Target-volume (cc)] None Recorded
[Target-packing-database-description] Rubber Mallet
[Target-xmin] 358
[Target-xmax] 443
[Target-ymin] 227
[Target-ymax] 374
[Target-zmin] 136
[Target-zmax] 191
[Target-level-of-difficulty] low
[Target-location-code] cac
[Target-orientation-code] z-

[Number-of-pseudo-targets] 1

[Pseudo-target-num] 1
[Pseudo-target-id] 6026
[Label-pixels] 161827
[Label-volume (cc)] 209.0
[Label-column-first] 217
[Label-column-last] 309
[Label-row-first] 306
[Label-row-last] 373
[Label-slice-first] 121
[Label-slice-last] 168
[Label-dimension-row (mm)] 62.2
[Label-dimension-col (mm)] 85.4
[Label-dimension-slice (mm)] 70.5
[Pseudo-target-material-form] bulk
[Pseudo-target-material-subtype] powder
[Pseudo-target-dimension-x (mm)] None Recorded
[Pseudo-target-dimension-y (mm)] None Recorded
[Pseudo-target-dimension-z (mm)] None Recorded
[Pseudo-target-mass (g)] 277.0
[Pseudo-target-volume (cc)] 250.0
[Pseudo-target-object-database-description] TA_MH01 plastic bottle + powder
[Pseudo-target-xmin] 215

[Pseduo-target-xmax] 309
[Pseduo-target-ymin] 305
[Pseduo-target-ymax] 371
[Pseduo-target-zmin] 121
[Pseduo-target-zmax] 168
[Pseduo-target-level-of-difficulty] high
[Pseduo-target-location-code] bab
[Pseduo-target-orientation-code] x-
[Pseduo-target-packing-database-description] P, TA_MH01

Information for ATR label image

[Number-of-labels] 6

[Label-num] 1
[Label-id] 1
[Label-pixels] 377501
[Label-volume (cc)] 487.6
[Label-column-first] 90
[Label-column-last] 309
[Label-row-first] 334
[Label-row-last] 370
[Label-slice-first] 56
[Label-slice-last] 127
[Label-dimension-row (mm)] 33.4
[Label-dimension-col (mm)] 203.2
[Label-dimension-slice (mm)] 106.5

[Label-num] 2
[Label-id] 2
[Label-pixels] 103484
[Label-volume (cc)] 133.7
[Label-column-first] 252
[Label-column-last] 407
[Label-row-first] 260
[Label-row-last] 312
[Label-slice-first] 103
[Label-slice-last] 145
[Label-dimension-row (mm)] 48.3
[Label-dimension-col (mm)] 143.8
[Label-dimension-slice (mm)] 63.0

[Label-num] 3
[Label-id] 3
[Label-pixels] 64612
[Label-volume (cc)] 83.5
[Label-column-first] 219
[Label-column-last] 307

[Label-row-first] 308
[Label-row-last] 369
[Label-slice-first] 123
[Label-slice-last] 166
[Label-dimension-row (mm)] 56.6
[Label-dimension-col (mm)] 81.7
[Label-dimension-slice (mm)] 64.5

[Label-num] 4
[Label-id] 4
[Label-pixels] 111611
[Label-volume (cc)] 144.2
[Label-column-first] 361
[Label-column-last] 438
[Label-row-first] 241
[Label-row-last] 369
[Label-slice-first] 142
[Label-slice-last] 187
[Label-dimension-row (mm)] 118.8
[Label-dimension-col (mm)] 71.5
[Label-dimension-slice (mm)] 67.5

[Label-num] 5
[Label-id] 5
[Label-pixels] 42787
[Label-volume (cc)] 55.3
[Label-column-first] 117
[Label-column-last] 252
[Label-row-first] 363
[Label-row-last] 390
[Label-slice-first] 202
[Label-slice-last] 286
[Label-dimension-row (mm)] 25.1
[Label-dimension-col (mm)] 125.3
[Label-dimension-slice (mm)] 126.0

[Label-num] 6
[Label-id] 6
[Label-pixels] 122608
[Label-volume (cc)] 158.4
[Label-column-first] 160
[Label-column-last] 351
[Label-row-first] 307
[Label-row-last] 335
[Label-slice-first] 209
[Label-slice-last] 367
[Label-dimension-row (mm)] 26.0
[Label-dimension-col (mm)] 177.2

[Label-dimension-slice (mm)] 237.0

Score Summary

[Num-target-detections] 1
[Num-target-misses] 1
[Num-pseudo-target-detections] 0
[Num-pseudo-target-misses] 1
[Num-false-alarms] 5
[Num-incomplete-detections] 2

Detections (Targets)

[Detection-number] 1 of 1
[Target-id] 6002
[Target-material-form] bulk
[Target-material-subtype] saline
[ATR-label-id-number] 1 of 1
[ATR-label-id] 2
[Precision] 1.00
[Recall] 0.52

Detections (Pseudo-targets)

NONE

False Alarms

NOTE: Intersecting GT labels include both targets AND pseudo-targets.

NOTE: Intersecting GT labels reported only if alpha*p, alpha*r is met, AND GT and ATR labels intersect by at least one pixel.

[False-alarm-number] 1 of 5
[ATR-label-id] 6
[Num-intersecting-gt-labels] 0

[False-alarm-number] 2 of 5
[ATR-label-id] 5
[Num-intersecting-gt-labels] 0

[False-alarm-number] 3 of 5
[ATR-label-id] 4
[Num-intersecting-gt-labels] 1
[Intersecting-target-id-number] 1 of 1
[Intersecting-target-id] 6004

[Target-material-form] bulk
[Target-material-subtype] rubber
[Precision] 1.00
[Recall] 0.29

[False-alarm-number] 4 of 5
[ATR-label-id] 3
[Num-intersecting-gt-labels] 1
[Intersecting-target-id-number] 1 of 1
[Intersecting-target-id] 6026
[Target-material-form] bulk
[Target-material-subtype] powder
[Precision] 1.00
[Recall] 0.40

[False-alarm-number] 5 of 5
[ATR-label-id] 1
[Num-intersecting-gt-labels] 0

Misses (Targets)

[Miss-number] 1 of 1
[Target-id] 6004
[Target-material-form] bulk
[Target-material-subtype] rubber
[Num-intersecting-atr-labels] 2
[Intersecting-atr-label-id-number] 1 of 2
[Intersecting-atr-label-id] 2
[Precision] 0.00
[Recall] 0.00
[Intersecting-atr-label-id-number] 2 of 2
[Intersecting-atr-label-id] 4
[Precision] 1.00
[Recall] 0.29

Misses (Pseudo-targets)

[Miss-number] 1 of 1
[Pseudo-target-id] 6026
[Pseudo-target-material-form] bulk
[Pseudo-target-material-subtype] powder
[Num-intersecting-atr-labels] 1
[Intersecting-atr-label-id-number] 1 of 1
[Intersecting-atr-label-id] 3
[Precision] 1.00
[Recall] 0.40

Appendix C: Example dder False Alarms Log File

SSN	ATR_Label_ID	Intersecting_Target_ID	Intersecting_Target_Form	Intersecting_Target_Subtype	Precision	Recall
8	6	NA	NA	NA	NA	NA
8	5	NA	NA	NA	NA	NA
8	4	6004	bulk	rubber	1	0.29
8	3	6026	bulk	powder	1	0.4
8	1	NA	NA	NA	NA	NA

Appendix D: Example pdpfa Summary Log File

[Program-name] pdpfa
[Version] 1.4
[Date] 05/10/14
[Time] 12:10:06

Command Line Information

[Input-log-list-filename] /home/franco/to4/logs/v35/gen_pdpfa_list.txt
[Object-database-filename] /home/franco/to4/to4-tools/dbase/odb.csv
[Packing-database-filename] /home/franco/to4/to4-tools/dbase/pdb.csv
[Summary-log-filename] /home/franco/to4/logs/v35/pdpfa_logs/pdpfa_log_summary.txt
[Detection-log-filename] /home/franco/to4/logs/v35/pdpfa_logs/pdpfa_log_detections.xls
[False-alarm-log-filename] /home/franco/to4/logs/v35/pdpfa_logs/pdpfa_log_false_alarms.xls

NOTE: See [Input-log-list-filename], [Detection-log-filename], and/or [False-alarm-log-filename] files for ist of SSNs used

NOTE: PD = (# detections)/(# targets)
NOTE: PFA = (total # false alarms)/(total # non-targets)
NOTE: Average number false alarms = (total # false alarms)/(total # scans with at least one false alarm)

[Total-num-scans] 188
[Total-num-objects] 1851
[Total-num-non-targets] 1366
[Total-num-targets-and-pseudo-targets] 485
[Total-num-targets] 412
[Total-num-pseduo-targets] 73

[PFA] 0.24
[Average-num-false-alarms] 2.36

PD for targets only

[PD-targets-overall] 0.76

[PD-targets-low-difficulty] 0.79
[PD-targets-high-difficulty] 0.74

[PD-targets-clay] 0.78
[PD-targets-rubber] 0.86
[PD-targets-saline] 0.62

[PD-targets-bulk] 0.69

[PD-targets-sheet] 0.89

PD for pseudo-targets only

[PD-pseudo-targets-overall] 0.34
[PD-pseudo-targets-low-difficulty] N/A
[PD-pseudo-targets-high-difficulty] 0.34

[PD-pseudo-targets-clay] 0.90
[PD-pseudo-targets-rubber] 0.30
[PD-pseudo-targets-saline] 0.63
[PD-pseudo-targets-powder] 0.03

[PD-pseudo-targets-bulk] 0.35
[PD-pseudo-targets-sheet] 0.30

PD for targets AND pseudo-targets

[PD-targets-and-pseudo-targets-overall] 0.69

[PD-targets-and-pseudo-targets-low-difficulty] 0.79
[PD-targets-and-pseudo-targets-high-difficulty] 0.66

[PD-targets-and-pseudo-targets-clay] 0.79
[PD-targets-and-pseudo-targets-rubber] 0.83
[PD-targets-and-pseudo-targets-saline] 0.62
[PD-targets-and-pseudo-targets-powder] 0.03

[PD-targets-and-pseudo-targets-bulk] 0.63
[PD-targets-and-pseudo-targets-sheet] 0.85

Appendix E: Example pdpfa Detections Log File

NOTE: Fields for which there is no information recorded in the object database are marked "None Reported"

SSN	Target ID	Detected	Material Type	Material Subtype	Material Form	Difficulty	Mass [g]	Volume [cc]	Dim x [mm]	Dim y [mm]	Dim z [mm]	Bbox x-min	Bbox x-max	Bbox y-min	Bbox y-max	Bbox z-min	Bbox z-max	Object db description	Packing db description
4	6004	0	t	rubber	bulk	low	1025	None Recorded	None Recorded	None Recorded	None Recorded	317	410	198	351	25	78	Rubber Mallet	Rubber mallet
4	6002	1	t	saline	bulk	low	253	None Recorded	None Recorded	None Recorded	None Recorded	126	200	243	343	188	231	Breast Milk Bottle 5% Saline	Saline
5	6004	0	t	rubber	bulk	low	1025	None Recorded	None Recorded	None Recorded	None Recorded	127	211	201	349	210	263	Rubber Mallet	Rubber Mallet
5	6002	0	t	saline	bulk	low	253	None Recorded	None Recorded	None Recorded	None Recorded	209	280	246	346	224	267	Breast Milk Bottle 5% Saline	saline
6	7006	0	t	saline	bulk	high	0	None Recorded	None Recorded	None Recorded	None Recorded	179	338	271	341	40	142	Merged - saline in bottle and saline in bag	(merged) saline

SSN	Target ID	Detected	Material Type	Material Subtype	Material Form	Difficulty	Mass [g]	Volume [cc]	Dim x [mm]	Dim y [mm]	Dim z [mm]	Bbox x-min	Bbox x-max	Bbox y-min	Bbox y-max	Bbox z-min	Bbox z-max	Object db description	Packing db description
6	7007	1	t	rubber	sheet	high	0	0	None Recorded	None Recorded	None Recorded	115	388	170	353	30	244	Merged - two rubber sheets	(merged) rubber sheet
7	7007	1	t	rubber	sheet	high	0	0	None Recorded	None Recorded	None Recorded	110	396	231	351	29	241	Merged - two rubber sheets	(merged) rubber sheet
7	6011	0	t	saline	bulk	high	285	0	None Recorded	None Recorded	None Recorded	148	217	276	328	29	124	Breast Milk bottle 10% Saline	saline
7	6012	0	t	saline	bulk	high	285	0	0	0	0	137	239	227	288	27	130	Breast milk bag 10% Saline	saline

Appendix F: Example pdpfa PD Log File

Target Type	Target Subtype or Form	Level of Difficulty	Num Targets	Num Detected	PD [%]
All	All	All	485	337	69.5
All	Clay	All	121	96	79.3
All	Rubber	All	173	143	82.7
All	Saline	All	157	97	61.8
All	Powder	All	34	1	2.9
All	Bulk	All	338	212	62.7
All	Sheet	All	147	125	85
All	All	High	376	248	66
All	Clay	High	92	71	77.2
All	Rubber	High	134	109	81.3
All	Saline	High	116	67	57.8
All	Powder	High	34	1	2.9
All	Bulk	High	254	147	57.9
All	Sheet	High	122	101	82.8
Target	All	All	412	312	75.7
Target	Clay	All	111	87	78.4
Target	Rubber	All	163	140	85.9
Target	Saline	All	138	85	61.6
Target	Bulk	All	275	190	69.1
Target	Sheet	All	137	122	89.1
Target	All	Low	96	76	79.2
Target	Clay	Low	29	25	86.2
Target	Rubber	Low	28	23	82.1
Target	Saline	Low	39	28	71.8
Target	Bulk	Low	71	52	73.2
Target	Sheet	Low	25	24	96
Target	All	High	303	223	73.6
Target	Clay	High	82	62	75.6
Target	Rubber	High	124	106	85.5
Target	Saline	High	97	55	56.7
Target	Bulk	High	191	125	65.4
Target	Sheet	High	112	98	87.5
Pseudo-target	All	High	73	25	34.2
Pseudo-target	Clay	High	10	9	90
Pseudo-target	Rubber	High	10	3	30
Pseudo-target	Saline	High	19	12	63.2

Target Type	Target Subtype or Form	Level of Difficulty	Num Targets	Num Detected	PD [%]
Pseudo-target	Powder	High	34	1	2.9
Pseudo-target	Bulk	High	63	22	34.9
Pseudo-target	Sheet	High	10	3	30

Num Non-targets	Num FAs	PFA [%]
1366	330	24.2

Num FAs	Num Scans with FAs	Avg Num FAs
330	140	2.36

Appendix G: Example pdpfa False Alarms Log File

SSN	ATR Label ID	Intersecting Target ID	Intersecting Target Form	Intersecting Target Subtype	Intersecting Target Precision	Intersecting Target Recall
4	1	NA	NA	NA	NA	NA
5	1	6004	bulk	rubber	1	0.28
6	4	NA	NA	NA	NA	NA
6	3	7006	bulk	saline	1	0.39
6	2	7006	bulk	saline	0.95	0.31
7	3	NA	NA	NA	NA	NA
7	2	NA	NA	NA	NA	NA

Appendix H: Example of gen_pdpfa.sh Summary Log File

[Program-name] gen_pdpfa.sh
[Timestamp] Mon Mar 24 20:58:48 CDT 2014

Command Line Information

[CT-dir] /home/franco/to4/ct
[GT-dir] /home/franco/to4/gt
[Labels-dir] /home/franco/to4/labels
[Logs-dir] /home/franco/to4/logs
[Label-image-prefix] A
[ATR-binary] /home/franco/to4/to4-tools/satr/satr
[SSN-list-filename]

NOTE: The remainder of this logfile is output from the atr, *dder*, and *pdpfa* programs

Appendix I: Example of gtver Log File

[Program-name] gtver
[Version] 1.2
[Date] 05/10/14
[Time] 11:41:22

Command Line Information

[Input-CT-image-filename]: /home/franco/to4/ct/I004.fits.gz
[Input-GT-image-filename]: /home/franco/to4/gt/to_verify/G004.fits.gz
[Input-odb-filename]: /home/franco/to4/to4-tools/dbase/odb.csv
[Input-pdb-filename]: /home/franco/to4/to4-tools/dbase/pdb.csv
[Output-log-filename]: /home/franco/to4/gt/gtver_logs/gtver_log_004.txt

Verification Output

NOTE: This program verifies targets AND pseudo-targets

--- IMAGE SIZE VERIFICATION ---

CT image [nrow ncol nslice]: [512 512 304]
GT image [nrow ncol nslice]: [512 512 304]

--- IMAGE SIZE VERIFICATION: PASSED! ---

--- GT LABEL ID VERIFICATION ---

GT label IDs match target IDs from database for this scan.
Target/Label IDs for this image:
6004
6002

--- GT LABEL ID VERIFICATION: PASSED! ---

--- TARGET MASS VERIFICATION ---

Target/label ID: 6004
Mass from database [g]: 1025.0
Mass calculated [g]: 1069.7
Mass witin spec?: YES

Target/label ID: 6002
Mass from database [g]: 253.0

Mass calculated [g]: 265.6
Mass witin spec?: YES

--- TARGET MASS VERIFICATION: PASSED! ---

--- TARGET BOUNDING BOX VERIFICATION ---

Target/label ID: 6004
Calculated precision: 0.89
Calculated recall: 0.99
Precision/recall witin spec?: YES

Target/label ID: 6002
Calculated precision: 0.95
Calculated recall: 0.98
Precision/recall witin spec?: YES

--- TARGET BOUNDING BOX VERIFICATION: PASSED! ---

[Verification-summary]: PASSED

11.4.2 Ground Truth Labeling

"ALERT ATR Project: Ground Truth Labeling"

ALERT ATR Project: Ground Truth Labeling

Author: Franco Rupcich
Email: franco.rupcich@gmail.com
Phone: 414-559-3338

Version 4

Contents

1 Introduction

This document describes how ground truth labels were generated for the TO4 project. In general, these labels were created using MeVisLab, a public domain image processing program.

2 Definitions

See the document entitled "ALERT ATR Project: Top-Level Specifications" for the list of terms and acronyms for this project, as well as for details regarding targets versus. pseudo-targets, detection, false alarms, misses, probability of detection, probability of false alarm, recall, and precision.

3 Background

3.1 Ground Truth Labeling

The purpose of the ALERT ATR Project (also known as Task Order 4, or TO4) is to address improvements in CT-based explosive detection equipment by developing improved ATR algorithms. An ATR algorithm takes as input a CT image containing objects, some of which are targets, and outputs a label image, which contains labels of objects for which the ATR has considered as targets. In order to assess the performance of an ATR algorithm, the label image it produces must be scored against a corresponding label image containing all targets in the corresponding CT image. This type of label image is referred to as the “ground truth label image.”

Ground truth labeling requires that for each target in a 2D CT slice all pixels in that slice for that target must be identified and given a unique label. All the labels for all the slices that contain the target are combined into a single image, called the *ground truth label image*, which is the same size (x,y, and z) as the original CT image. This process is denoted segmentation. In general, the segmentation can be performed using one of the three following methods:

1. Automatic segmentation
2. Manual segmentation
3. Semi-automated segmentation

The first method is not a viable option, since no known automated segmentation method is able to provide sufficient performance in segmenting any given object type from airline luggage. This is due to a number of reasons, including splitting of objects as a result of image artifacts (beam-hardening, metal streak and shading artifacts) and merging of physically touching objects with similar CT numbers. Indeed, improving automated segmentation techniques is a major goal of TO4.

The second method requires manual segmentation of each target in each slice of an image, making this approach prohibitively time-consuming or prohibitively expensive.

The third method is the middle ground between the first two. Semi-automated techniques allow the human user to segment the objects known to be targets. The computer-assisted aspect of the tool can speed up the segmentation process by guiding the contours being drawn by the user and by interpolating contours for an object across slices.

3.2 Selecting a Ground Truth Labeling Tool

Several free visualization and processing software packages [1][2][3] were investigated by Seemeen Karimi during Task Order 1: Segmentation Initiative (TO1), which also required that ground truth labeling be performed [4]. Karimi reported that while these tools are simple to learn, they do not allow interpolation between contours, a significant time-saving process. In addition, Karimi concluded that some of these programs offer other methods for semi-automatic segmentation, but the results cannot be manually edited.

Karimi chose to use the MeVisLab image processing software tools for ground truth labeling during TO1, as they allowed for guided contouring of images, as well as for interpolation of contours across image slices.

For this project, we chose to use MeVisLab based on Karimi’s recommendations. In addition, we were able to modify the existing code that Karimi had created for TO1 to suit the purposes of TO4.

3.3 MeVisLab

The MeVisLab image processing software provides an image viewer and image processing tools that enable computer-assisted segmentation. MeVisLab is being developed and used by MeVis Medical Solutions AG and Fraunhofer MEVIS (formerly MeVis Research GmbH) in Bremen, Germany.

An image processing task is achieved by creating a graphical program in MeVisLab. The graphical program is also called a network. The network is created of interconnected modules. Each module performs a specific operation. A module may receive input data from a module and may create output data that it passes to the next module. A module may provide control or book-keeping functions.

The MeVisLab software can be downloaded from the following website.

http://www.mevislab.de/download/

The software runs on Windows, OS X, and Linux operating systems. The Mac OS X version was used in this work.

The MeVisLab forums provide additional support from the developers and can be found at

http://forum.mevis.fraunhofer.de/index.php

4 Ground Truth Labeling using MeVisLab

4.1 Overview

We have created a computer-assisted manual segmentation network, called *TO4 Ground Truth Segment and Labeling Network.mlab*, which is a modified version of Karimi's network created during TO1. The following sections outline the basic steps taken to create a ground truth image. Detailed step-by-step instructions are given in Section 4.2. Figure 1 is an image of the network in MeVisLab.

4.1.1 Load CT Image

The user must first load a DICOM version of a CT image containing the objects that will be labeled. A label image is created using the same dimensions as the CT image. To start, all pixel values of the label image are zero. Individual labels will subsequently be inserted into the label image to create the final GT label image containing all labels for all contoured objects from the given CT image.

4.1.2 Segmentation (Contouring)

The network assists the operator in drawing contours for an object using underlying semi-automated segmentation techniques. A contour in MeVisLab is called a Contour Segmentation Object (CSO). Lists of CSOs for a given image can be saved and loaded. Once the user has created contours for a given object on several slices, the contours can be interpolated across those slices. The interpolation of contours across slices is the single largest time-saving step because it allows a 3D volume to be segmented by drawing contours on only a fraction of the slices as opposed to each slice. MeVisLab's ability to assist in drawing contours and interpolate contours on any of the three standard multi-planar reconstruction planes (transversal, sagittal, and coronal) is one of the primary reasons it was chosen as the software tool to perform ground truth labeling.

4.1.3 Labeling

All points within the 3D contoured volume can be filled in with a single label (mask) value. The label values used were the target IDs specified in the object/packing database. In the case of multiple objects of interest, the individual objects can be separately labeled and combined together into the ground truth label image.

4.1.4 Region Growing

Region growing may be used to make the label more accurate. For example, if an object contains a hole in the center, it may be easier to contour and label the entire object (including the hole) and then use region growing with appropriate threshold values such that the hole is not included in the label.

4.1.5 Add Individual Labels to Label Image

Once a label has been created for a single object, it can be inserted to the label image. The user then continues to label additional objects and add them to the label image.

4.1.6 Save GT Label Image

Once all desired objects have been labeled and added to the label image, the user saves the label image in .raw format (16-bit unsigned short). Images can then be converted from .raw to .fits format using the *raw2fits* conversion tool distributed as part of the TO4 Software Tools (see the document entitled "TO4 Software Tools Spec"), and gzipped using *gzip* to produce a fits.gz format.

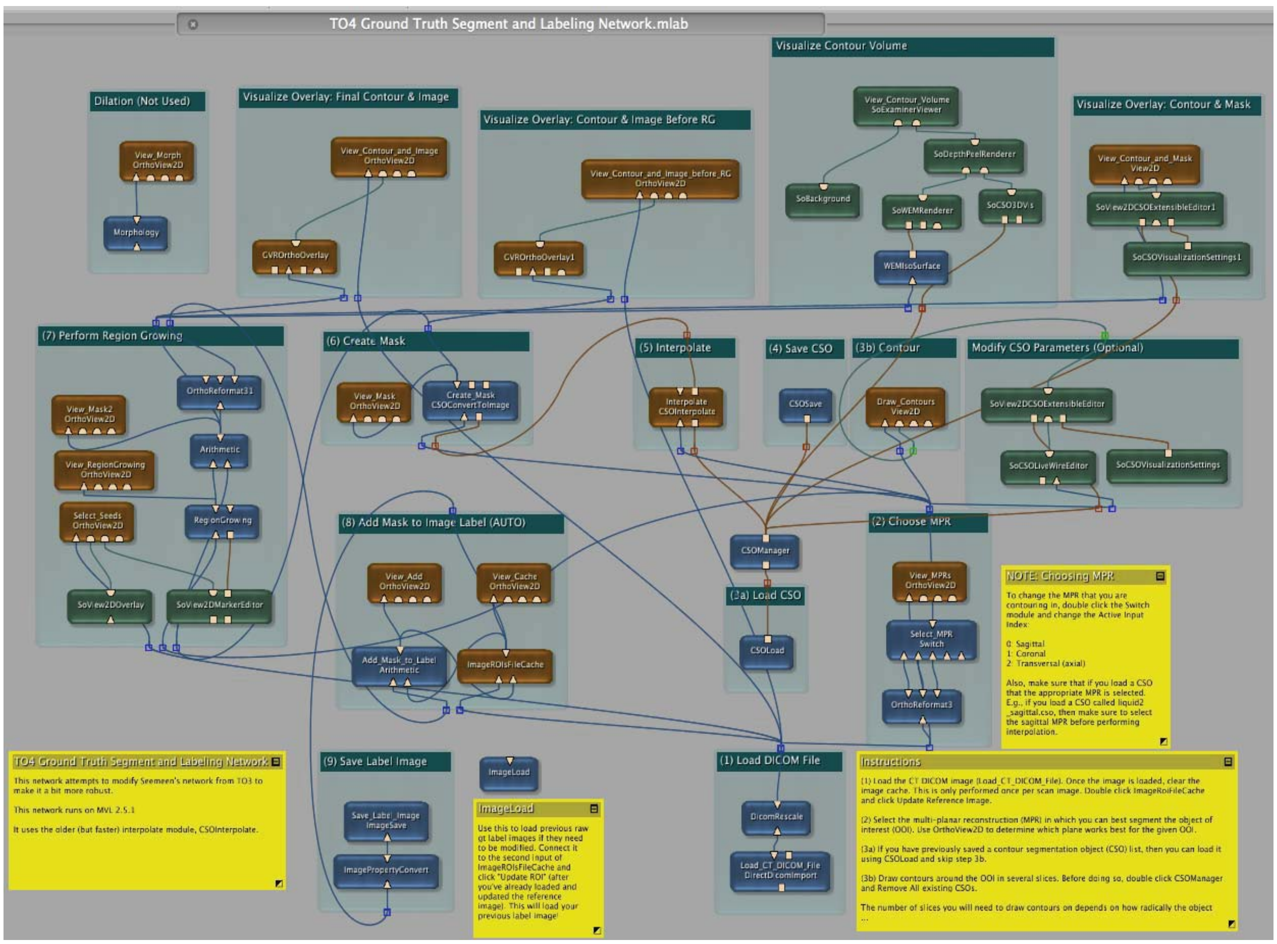


Figure 1: TO4 Ground Truth Segment and Labeling Network

4.2 Instructions for Use

NOTE: The step numbers of these instructions coincide with the numbers of the groups with named headings in the network (see Figure 1).

1. Load a CT DICOM image using the "Load_CT_DICOM_File" module. To load the CT image, browse to the location of the file and select OK. Then click the "Clear Log + Import" button.

   WARNING: If trying to load an image comprising multiple single-slice DICOM files, MeVisLab seems to load the image in reverse order, i.e. the first slice is the last slice. To avoid this, load the image as a single DICOM file containing all slices of the image. To convert multiple single-slice DICOM files to a single multi-slice DICOM file, use the Tudor DICOM package [found here: http://santec.tudor.lu/project/dicom] for ImageJ.

   Double click the "ImageROIsFileCache" and click "Update Reference Image." This creates a zeroed out label image with the same dimensions as the input CT image. Labels will be added to this image, and the final aggregated label image will eventually be saved as the GT label image.

   NOTE: Updating the Reference Image is only done the first time a CT image is loaded. Clicking "Update Reference Image" after labels have been added will zero out the image.

2. Double click the "Select_MPR" module to select the MPR that will provide the most straightforward contouring of the OOI (0=Sagittal, 1=Coronal, 2=Transversal). To view the image in different MPRs, double click the "View_MPRs" module, which will bring up an image-viewing pane, and then use the drop-down box in the upper left corner of the window to select a different MPR.

   NOTE: You can change the window/level of the image by right-clicking on the "View_MPRs" module and selecting "Show Window->Settings," or you may use the right mouse-button on the image itself.

3. If you have previously saved a contour segmentation object (CSO) list, then you can load it using the "CSOLoad" module and skip to step 4.

   Otherwise, double click the "Draw_Contours" module to bring up the image. Draw contours around the OOI in several slices. Each contour is saved as a CSO and can be viewed and managed by double clicking the "CSOManager" module.

   NOTE: Before beginning to draw contours for a new OOI, double click "CSOManager" and click "Remove All" to remove any CSOs from a previous OOI.

   NOTE: The number of slices you will need to draw contours on depends on how radically the object changes from slice to slice. Generally you can contour every 5-20 slices.

   NOTE: You can change the window/level of the image by right-clicking on the "Draw_Contours" module and selecting "Show Window->Automatic Panel," or you may use the right mouse-button on the image itself.

4. Save the CSO list for this object using the "CSOSave" module. This is done so that the CSO list may be loaded later without having to re-contour the image (i.e., in case MeVisLab crashes, or in case the image needs to be re-labeled).

5. Interpolate the contours by double clicking the "Interpolate" module and clicking "Apply."

   WARNING: Every so often, this step will cause MeVisLab to crash. From my experience, it will crash if
   1. not enough contours have been drawn before interpolation, especially when the object changes radically (for example, at the beginning and end of an object, especially sheets)

   2. contours are drawn poorly before interpolation (for example, a single contour that overlaps itself)

   It is best to draw a few contours and then click "Interpolate." If MeVisLab does not crash, you can save the CSO list using the "CSOSave" module and continue drawing contours. If at some later point during contouring of the OOI MeVisLab does crash, you now have a saved CSO list that you can load in step 3 instead of having to start over with the contouring.

   TIP: For a single object with disparate components within a single slice, contour and interpolate the two components separately, but use the same mask value. This is to prevent the interpolator from trying to interpolate multiple components across slices.

6. Double click the "Create_Mask" module to create the image mask (label). The "Foreground Value" field dictates the value of the label. Click "Update" to create the label.

   NOTE: You can check how well the pre-region-growing label matches up with the OOI in the CT image using View_Contour_and_Image_before_RG. In addition, View_Contour_and_Mask allows you to determine how well the label lines up with the drawn contours, and View_Contour_Volume allows you to view the interpolated contours in 3D volumetric form.

7. Use the "Select_Seeds" module to select one or more seed points in the OOI for region growing. Manually select lower and upper thresholds for the region growing algorithm using the "RegionGrowing" module.

   NOTE: You can check how well the label matches up with the OOI in the CT image using View_Contour_and_Image. If you're not satisfied with the label, you can select different region growing thresholds, or you can start at step 3 and try drawing additional contours.

8. Add the current label to the aggregated label image by double clicking the "ImageROIsFileCache" module and then clicking "Update Roi." Once the image has been added, double click the "SoView2DMarkerEditor" module and click "Delete All." This will remove all seed points in preparation for the next OOI to be contoured.

   NOTE: You can view the aggregated label image by double clicking the "View_Cache" module.

REPEAT STEPS 2-8 FOR EACH OOI IN THE CT IMAGE.

9. Save the aggregated label image using the "Save_Label_Image" module.

5 Lessons Learned about applying MVL

5.1 Online Support

MeVisLab does not have very extensive documentation, and so its learning curve is quite steep, especially to those with little computer programming and/or image processing experience. However, the community of users and developers that participate in the MeVisLab developer's forum (http://forum.mevis.fraunhofer.de/index.php) has proven to be very knowledgeable and quick to help with any issue.

5.2 Crashing during interpolation

MeVisLab is not a very stable program, and it may crash from time to time. Through extensive use of the program, it was empirically determined that the primary causes of the program crashing are

1. Not enough contours have been drawn before interpolation, especially when the object changes radically (for example, at the beginning and end of an object, especially sheets)

2. Contours are drawn poorly before interpolation (for example, a single contour that overlaps itself)

The **WARNING** in step 5 in the instructions reiterates this point and indicates that the user should save the CSO list often so as not to lose their work if the program does crash.

5.3 Thin Sheets

Thin sheets (approximately <5 pixels, or 5mm, thick) are very difficult to segment in MeVisLab. This is due in part to the active contouring algorithm sometimes jumping over the sheet to the next nearest object. The presence of CT artifacts (streaking and shading) further complicates the process of segmenting thin sheets. Further, MeVisLab will sometimes crash during segmentation of these thinner objects (see Section 5.2). The best approach, though time-consuming, is to segment many slices (every 5-10) when segmenting thin sheets.

5.4 CT Artifacts

The presence of metal in many of the bags causes severe streaking/shading artifacts. These artifacts can cause splitting of an object, and even cause portions object to become invisible. In these cases, region growing is of little use (since the object will have very high and very low CT numbers from the artifacts). Thus, the user should segment the object in as many slices as possible, especially those containing artifacts and take care to segment as close to the object as possible.

6 References

[1] NIH, "itk-SNAP," http://www.itksnap.org/pmwiki/pmwiki.php

[2] NIH, "Image J Image Processing and Analysis in Java," http://rsbweb.nih.gov/ij/

[3] Philippe PUECH and Loic BOUSSEL, "DICOMWorks," http://dicom.online.fr/

[4] Karimi, Seemeen. "Computer-Assissted Manual Segmentation Method for Grand Challenge," May 18, 2010.

7 Revision History

Version	Changes
1	Initial revision
2	Modified based on feedback from Carl Crawford. Added more background. Added Lessons Learned section.
3	Added additional background and lessons learned.
4	Minor revision based on feedback from Carl Crawford.

11.4.3 Simulated Test Images

"ALERT ATR Project: Simulated Test Images Specification"

ALERT ATR Project: Simulated Test Images Specification

Author: Franco Rupcich
Email: franco.rupcich@gmail.com
Phone: 414-559-3338

Version 2

Contents

1 Introduction

This document provides details about the simulated test images generated for the purpose of testing ATR algorithms for the TO4 project.

2 Acronyms

Term	Definition
ATR	Automated Target Recognition
CT	Computerized tomography
g3d	Program used to simulate 3D images of simple shapes
GT	Ground truth
TO4	Task Order 4. This project: the ATR Initiative

3 Simulated Test Images

Two simulated CT test images and their corresponding GT label images are provided. The images were generated using g3d [1], a program that uses parameters specified in shape files to create 3D images of simple shapes (e.g., rectangles, cylinders, etc). Images generated from g3d were converted from mi format to 3D FITS (16-bit unsigned short) format using *mi2fits* [2] and compressed using *gzip*.

A sample 2D slice of each simulated CT and GT label image is shown in Figure 1 and Figure 2. Table 1 describes the simulated CT image properties and contents. Note that the GT label images have the same name as their corresponding CT images except with the prefix "G" instead of "I." Table 2-4 describe the properties of the shapes in the simulated images, as well as the GT label ID for each target. All x, y, z, length, and radius values are in pixels. The thickness of the rubber sheet targets is calculated assuming 1 pixel is equal to 0.928mm, the same slice thickness as in the TO4 CT dataset.

Table 1: Simulated CT image properties and contents

CT Image Name	Image Dimensions [x y z]	# Objects	# Targets	Contents
I200.fits.gz	[512 512 281]	3	2	1) Saline bulk target 2) Clay bulk target 3) Water non-target
I201.fits.gz	[512 512 281]	3	2	1) Rubber sheet target 2) Rubber sheet target 3) Water non-target

Table 2: Shape properties for I200 and G200

Shape	x0	y0	z0	x1	y1	z1	radius	Simulated Object	Image Pixel Value [MHU]	GT Label ID
Cylinder 1	0	0	-50	0	0	50	30	Saline	1080	6002
Cylinder 2	100	-50	0	100	50	0	30	Clay	1640	6051
Cylinder 3	-50	100	0	50	-100	0	30	Water	1000	NA

ALERT ATR Project: Simulated Test Images, Page 4

Table 3: Shape properties for rectangles in I201 and G201

Shape	x0	x-length	y0	y-length	z0	z-length	Simulated Object	Image Pixel Value [MHU]	GT Label ID
Rectangle 1	-30	200	100	8	-20	100	Rubber sheet (0.34")	1361	6053
Rectangle 2	-30	7	-20	100	0	75	Rubber sheet (0.30")	1252	6144

Table 4: Shape properties for cylinder in I201 and G201

Shape	x0	y0	z0	x1	y1	z1	radius	Simulated Object	Image Pixel Value [MHU]	GT Label ID
Cylinder 1	150	-50	0	100	50	0	30	Water	1002	NA

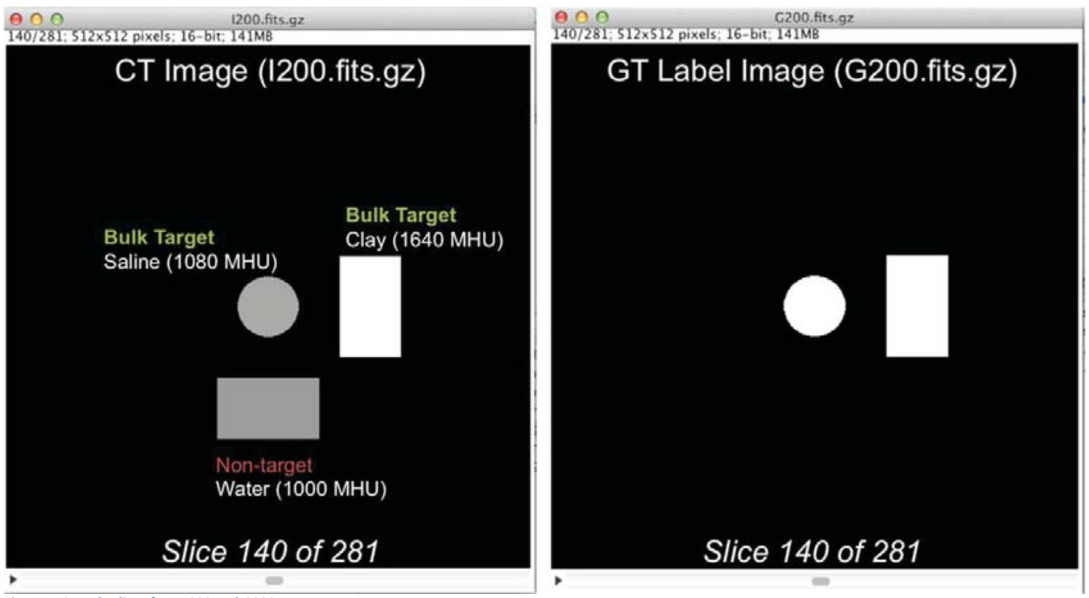


Figure 1: Sample slices from I200 and G200

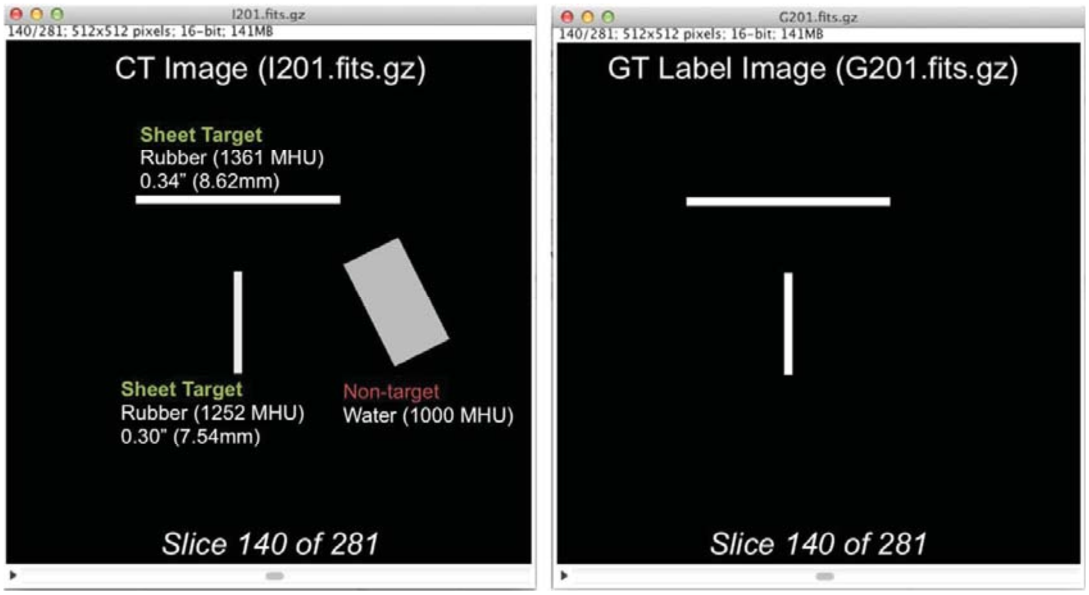


Figure 2: Sample slices from I201 and G201

4 References

[1] Crawford, C. R., and King, K. F., "Computed tomography scanning with simultaneous patient translation," Medical Physics, Vol. 17, No. 6, November/December 1990, pp. 967-982.

[2] ALERT ATR Project: Software Tools Specification

[3] ALERT ATR Project: Top Level Specification

5 Revision History

Version	Changes
1	Initial revision
2	Added tables describing shape properties

11.5 Program Review

11.5.1 Agenda+

11.5.1.1 Agenda

"ATR Development for CT-Based EDS (Task Order 4) Program Review Agenda"

ALERT
AWARENESS AND LOCALIZATION OF EXPLOSIVES-RELATED THREATS

Awareness and Localization of Explosives-Related Threats (ALERT)

A Department of Homeland Security Center of Excellence

ATR Development for CT-Based EDS (Task Order 4) Program Review

November 6, 2014
Raytheon (Room 240), Egan Research Center, Northeastern University
120 Forsyth St., Boston, MA

AGENDA

Thursday, November 6, 2014

Time	Topic	Speaker	Affiliation
7:30 AM	**Check-in and Breakfast**		
	Introduction and Project Overview		
8:00 AM	Welcoming Remarks - ALERT	Michael Silevitch	ALERT / Northeastern University
8:05 AM	Welcoming Remarks - DHS	Laura Parker	Department of Homeland Security
8:10 AM	Project Overview	Carl Crawford	Csuptwo
8:40 AM	CT Datasets	Carl Crawford	Csuptwo
9:00 AM	Tools and Ground Truth	Franco Rupcich	Self
	ATR Development		
9:30 AM	ATR Development & Discussion	Dong Hye Ye Charlie Bouman Pengchong Jin	Purdue University
10:30 AM	**Break**		
11:00 AM	ATR Development & Discussion	Jens Gregor	University of Tennessee
12:00 PM	ATR Development	Jun Zhang Laura Drake Hongquan Zuo	University of Wisconsin, Milwaukee
1:00 PM	**Lunch**		
1:30 PM	ATR Development & Discussion	Synho Do	Massachusetts General Hospital
2:30 PM	ATR Development	Philip Top Ana Paula Sales Hyojin Kim Timo Bremer Steve Azevedo Harry Martz	Lawrence Livermore National Lab
3:30 PM	**Break**		

Awareness and Localization of Explosives-Related Threats (ALERT)

A Department of Homeland Security Center of Excellence

Time	Topic	Speaker	Affiliation
4:00 PM	Vendor Panel Discussion	David Lieblich	Analogic Corporation
		Kam Lin Wong	Reveal Imaging Technologies, Inc.
		David Perticone	L-3 Communications
		James Connelly	Integrated Defense and Security Solutions
		Carl Bosch	SureScan
		Ling Tang	Rapiscan Labratories, Inc.
		Piero Landolfi	Morpho Detection
		Christopher Gregory	Smiths Detection
	Discussion and Next Steps		
4:40 PM	Discussion, Next Steps	Carl Crawford	Csuptwo
		Clem Karl	Boston University
		Harry Martz	Lawrence Livermore National Lab
5:20 PM	Closing Remarks - DHS	Laura Parker	Department of Homeland Security
5:25 PM	Closing Remarks - ALERT	Michael Silevitch	Northeastern University / ALERT
5:30 PM	Adjourn	Carl Crawford	Csuptwo

11.5.1.2 Attendee List

"ATR Development for CT-Based EDS (Task Order 4) Program Review Attendee List"

Awareness and Localization of Explosives-Related Threats (ALERT)

A Department of Homeland Security Center of Excellence

ATR Development for CT-Based EDS (Task Order 4)
Program Review

November 6, 2014
Raytheon (Room 240), Egan Research Center, Northeastern University
120 Forsyth St., Boston, MA

ATTENDEE LIST

David Atkinson
Senior Research Scientist
Pacific Northwest National Laboratory

Erez Attias
Explosive Detection & Security Technology Branch
P.M.O

Stephen Azevedo
Explosives Detection Project Engineer
Lawrence Livermore National Laboratory

Kumar Babu
Branch Chief, Aviation Security
Department of Homeland Security

Sara Baier
Administrative Assistant
Northeastern University

Douglas Bauer
Visiting Scientist
University of Connecticut

John Beaty
Director of Technology Programs
Northeastern University

Moritz Beckmann
CEO
XinRay Systems LLC

Simon Bedford
Director of Government Programs
Astrophysics, Inc.

Deanna Beirne
Director of Computer Services
Northeastern University

Ralf Birken
Research Assistant Professor
Northeastern University

Carl Bosch
CTO
SureScan

Jeffrey Brasher
Sr. Project Manager
SRA International, Inc.

Emel Bulat
Industrial and Government Liaison
Northeastern University

Cindy Carey
IMS Applications Scientist
Bruker Detection Corporation

Frank Cartwright
Technology Portfolio Manager
TSA

David Castañón
Professor
Boston University

Gongyin Chen
Senior Scientist
Varian Medical Systems

Charles Choi
Lead Systems Engineer
General Dynamics

Allan Collier
Branch Chief, Air Cargo
TSA

Todd Combs
Director, Global Security Sciences Division
Argonne National Laboratory

James Connelly
CTO
Integrated Defense and Security Solutions

Sean Corrigan
Algorithm Engineer
SureScan

Andrew Cox
Systems Analyst
Sandia National Laboratories

Awareness and Localization of Explosives-Related Threats (ALERT)

A Department of Homeland Security Center of Excellence

Carl Crawford
President
Csuptwo, LLC

Synho Do
Assistant in Physics
Massachusetts General Hospital

Adam Erlich
Vice President, Marketing and Business Development
Triple Ring Technologies

Michael Finnin
Research Staff Member
Institute for Defense Analyses

Danny Fisher
Projects Manager
Israel Prime Minister's Office

Jonathan Foley
Tufts University

Erin Gallagher
Program Manager
Schafer Corporation

Stan German
Scientist
Charles River Analytics, Inc.

Galia Ghazi
graduate student
Northeastern University

Brian Gonzales
Simulation and Imaging Scientist
XinRay Systems LLC

Jens Gregor
Professor
University of Tennessee

Chris Gregory
Sr. Algorithm Developer
Smiths Detection

Adam Grosser
Director, Product Management and Business Development
Redlen Technologies, Inc.

Jeffrey Hamel
President
IDSS

Roy Hebert
President and CEO
RJH & Associates

Kristin Hicks
Director of Operations
Northeastern University

Matt Higger
Ph.D. Student
Northeastern University

Paul Hurd
Product Manager
Passport Systems, Inc.

Teri Incampo
Production and Reporting Specialist
Northeastern University

Sheldon Jacobson
University of Illinois

Edward Jimenez
Sandia National Laboratories

Clem Karl
Professor
Boston University

Vladimir Kekukh
Project Scientist
Bruker Corporation

Hyojin Kim
Lawrence Livermore National Laboratory

Margot Kimura
Senior Analyst
Sandia National Laboratories

Ronald Krauss
Bulk Explosives Detection Technology Lead
Department of Homeland Security

Pierfrancesco Landolfi
Director of Engineering, CT
Morpho Detection

Richard Lanza
Senior Scientist
Massachusetts Institute of Technology

David Lieblich
Director of Reconstruction and Detection
Analogic Corporation

Anne Magrath
Director of Finance and Research Contracts Administration Operations
Northeastern University

Awareness and Localization of Explosives-Related Threats (ALERT)

A Department of Homeland Security Center of Excellence

Limor Martin
Postdoctoral Researcher
Boston University

Harry Martz
Director
Lawrence Livermore National Laboratory

Patricia McDaniel
Combating Terrorism Technical Support Office and ManTech

Eric Miller
Professor and Chair
Tufts University

Andrey Mishin
Managing Partner
Radmedex, LLC

Richard Moore
Research Director of Breast Imaging
Massachusetts General Hospital

Edward Morrison
Professor & Head, Dept. of Anatomy, Physiology and Pharmacology
Auburn University

Edward Morton
Chief Technical Officer
Rapiscan Systems

Boris Oreper
Technical Director
L-3 Communications

Laura Parker
Program Manager
Department of Homeland Security

John Parmeter
Technical Staff Member
Sandia National Laboratories

Douglas Pearl
President
Inzight Consulting, LLC

David Perticone
Engineering Fellow
L-3 Communications

Patrick Radisson
CTO
Multixdetection

Paul Redfern
Aviation Security Research Programme Manager
U.K. Government

Yolanda Rodriguez-Vaqueiro
PhD Student
Northeastern University

Terri Rose
Program Manager
Oak Ridge National Laboratory

Susan Rose-Pehrsson
Director, Navy Technology Center for Safety and Survivability
Naval Research Laboratory

Masoud Rostami
Northeastern University

David Rundle
Sales and Applications R&D Engineer
Kromek

Franco Rupcich
Engineering Contractor
Self

Robert Shuchatowitz
Research Manager
Reveal Imaging Technologies, Inc.

Michael Silevitch
Robert D. Black Professor
Northeastern University

Melanie Smith
Partnership Events and Education Specialist
Northeastern University

Serge Soloviev
Principal Scientist
Reveal Imaging Technologies, Inc.

Samuel Song
CTO
TeleSecurity Sciences, Inc.

Jeff Stillson
Senior Director of Machine Vision
L-3 Communications

Simon Streltsov
Co-Founder and President
LongShortWay

Greg Struba
Senior Systems Engineer
Booz Allen Hamilton

Zachary Sun
Boston University

Awareness and Localization of Explosives-Related Threats (ALERT)

A Department of Homeland Security Center of Excellence

David Throckmorton
Air Cargo Program Manager
Department of Homeland Security

Jill Tomlinson-Phillips
Lead Scientist
Booz Allen Hamilton

Philip Top
Electrical Engineer
Lawrence Livermore National Laboratory

Alex Van Adzin
Vice President, Chief Legal Officer, and Corporation Secretary
Photo Diagnostic Systems, Inc.

Lou Wainwright
Sr. Director, Imaging, Instrumentation, X-ray
Triple Ring Technologies

Tim White
Scientist
Pacific Northwest National Laboratory

Alyssa White
Clinical Research Associate
Northeastern University

David Wiley
President and CTO
Stratovan Corporation

Cody Wilson
Technical Product Manager, Senior Scientist
Passport Systems, Inc.

Kam Lin Wong
Senior Algorithm Engineer
SAIC

William Worstell
Principal Research Scientist
Photo Diagnostic Systems, Inc.

Rich Wronski
Division Vice President
Charles River Analytics, Inc.

Yuxiang Xing
Associate Professor
Tsinghua Univiversity

Dong Hye Ye
Postdoctoral Researcher
Purdue University

Bridget Yu
Northeastern University

Jun Zhang
Professor
University of Wisconsin-Milwaukee

11.5.1.3 Speaker Bios

"ATR Development for CT-Based EDS (Task Order 4) Program Review Speaker Biographies"

Awareness and Localization of Explosives-Related Threats (ALERT)

A Department of Homeland Security Center of Excellence

ATR Development for CT-Based EDS (Task Order 4) Program Review

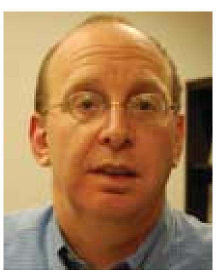
November 6, 2014
Raytheon (Room 240), Egan Research Center, Northeastern University
120 Forsyth St., Boston, MA

SPEAKER BIOGRAPHIES

Charles Bouman

Purdue University

Charles A. Bouman is the Showalter Professor of Electrical and Computer Engineering and Biomedical Engineering at Purdue University where he also serves has a co-director of Purdue's Magnetic Resonance Imaging Facility. He received his B.S.E.E. degree from the University of Pennsylvania, M.S. degree from the University of California at Berkeley, and Ph.D. from Princeton University in 1989. Professor Bouman's research focuses on inverse problems, stochastic modeling, and their application in a wide variety of imaging problems including tomographic reconstruction and image processing and rendering. Prof. Bouman is a Fellow of the IEEE, AIMBE, IS&T, and SPIE and is currently the IEEE Signal Processing Society's Vice President of Technical Directions. He has also served as the Editor-in-Chief of the IEEE Transactions on Image Processing and the Vice President of Publications for the IS&T Society.



Carl R. Crawford

Csuptwo, LLC

Carl R. Crawford, Ph.D., is president of Csuptwo, LLC, a technology development and consulting company in the fields of medical imaging and Homeland Security. He has been a technical innovator in the fields of computerized imaging for more than thirty years. Dr. Crawford was the Technical Vice President of Corporate Imaging Systems at Analogic Corporation, Peabody, Massachusetts, where he led the application of signal and image processing techniques for medical and security scanners. He developed the reconstruction and explosive detection algorithms for a computerized tomographic (CT) scanner deployed in airports worldwide. He was also employed at General Electric Medical Systems, Milwaukee, Wisconsin, where he invented the enabling technology for helical scanning for medical CT scanners, and at Elicit, Haifa, Israel, where he developed technology for cardiac CT scanners. He also has developed technology for magnetic resonance imaging (MRI), single photon emission tomography (SPECT), positron emission tomography (PET), ultrasound imaging (U/S), dual energy imaging and automated threat detection algorithms based on computer aided detection (CAD). Dr. Crawford has a doctorate in electrical engineering from Purdue University. He is a Fellow of the Institute of Electrical and Electronics Engineers (IEEE), is a Fellow of the American Association of Physicists in Medicine (AAPM), and is an associate editor of IEEE Transactions on Medical Imaging.

Awareness and Localization of Explosives-Related Threats (ALERT)

A Department of Homeland Security Center of Excellence

Laura Drake

University of Wisconsin, Milwaukee

Laura Drake received her B.A in mathematics from Johns Hopkins University, and M.S. and Ph.D. degrees in electrical engineering from Northwestern University in 2001. She has worked on satellite command and control software development at Space Applications Corporation in Sunnyvale, California, auditory and acoustic signal processing as a research assistant at Northwestern University and as a summer intern at AT&T Bell Labs, and image processing projects funded by SBIR research grants. Her interests are acoustic and image processing.

Synho Do

Massachusetts General Hospital

Synho Do, PhD, is an Assistant in Physics at Massachusetts General Hospital, where he is a technical committee member of Webster Center for Advanced Research and Education in Radiation, and Instructor at Harvard Medical School. Dr. Do received the Ph.D. degree in Biomedical Engineering from University of Southern California. He is currently a member of IEEE Signal Processing Society, Bio-Imaging and Signal Processing (BISP). He is a MGH site PI for nVidia CUDA Research Center (CRC). Dr. Do's current research interests include statistical signal and image processing, estimation, detection, and medical signal and image processing, such as computed tomography. He has been a Co-Investigator for multiple medical imaging projects, and Co-PI/PI on medical (i.e., GE, Siemens, and Philips etc) and security (i.e., DHS, DARPA etc) image reconstruction projects.

Jens Gregor

University of Tennessee

Dr. Jens Gregor received a PhD in Electrical Engineering from Aalborg University, Denmark in 1991. He then joined the Department of Computer Science at the University of Tennessee, Knoxville. Following a recent merger, he currently holds the rank of Professor in the Department of Electrical Engineering and Computer Science. His research spans the fields of pattern recognition, image reconstruction and parallel computing. This work has been published in a combined total of more than 80 book-chapters, journal articles and conference papers. He has developed and implemented statistical and algebraic imaging algorithms for medical and preclinical applications as well as waste management and non-destructive testing applications for several different data modalities. He was a participant in ALERT Task Order 3 which dealt with iterative reconstruction for luggage screening. He currently participates in ALERT Task Order 4 in regard to automated threat recognition. He has served as a consultant to Oak Ridge National Laboratory, Siemens Medical, Hexagon Metrology and various small companies in East Tennessee.

Awareness and Localization of Explosives-Related Threats (ALERT)

A Department of Homeland Security Center of Excellence

Harry Martz

Lawrence Livermore National Lab

Harry Martz is the Director for Non-destructive Characterization Institute (NCI) at Lawrence Livermore National Laboratory and PI on DHS S&T Explosive Division Explosive Detection Projects and DNDO Nuclear and Radiological Imaging Platform (NRIP) and Passive And X-ray Imaging Scanning (PAXIS). Harry joined the Laboratory in 1986 as a Physicist to develop the area of x-ray and proton energy loss computed tomography for the non-destructive inspection of materials, components, and assemblies. He received his M.S. and Ph.D. in Nuclear Physics/Inorganic Chemistry from Florida State University, and his B.S. in Chemistry from Siena Collage. Harry's interests include the research, development and application of nonintrusive characterization techniques as a three-dimensional imaging instrumentation to better understand material properties and inspection of components and assemblies, and generation of finite element models from characterization data. He has applied CT to inspect one-millimeter sized laser targets, automobile and aircraft components, reactor-fuel tubes, new production reactor target particles, high explosives, explosive shape charges, dinosaur eggs, concrete and for non-destructive radioactive assay of waste drum contents. Recent R&D efforts include CT imaging for conventional and homemade explosives detection in luggage and radiographic imaging of cargo to detect special nuclear materials and radiological dispersal devices. Dr. Martz has authored or co-authored over 300 papers and is co-author of a chapter on Radiology in Non-destructive Evaluation: Theory, Techniques and Applications, Image Data Analysis in Non-destructive Testing Handbook, third edition: Volume 4, Radiographic Testing, and contributed a chapter entitled Industrial Computed Tomographic Imaging to the Advanced Signal Processing Handbook: Theory and Implementation for Radar, Sonar and Medical Imaging Real-Time Systems. He has also served on several National Academy of Sciences Committees on Aviation Security and is the Chair of the Committee on Airport Passenger Screening: Backscatter X-Ray Machines. Harry has been co-chair of ALERT ADSA Workshops. Dr. Martz has presented a short course on CT imaging at The Center for Non-destructive Evaluation, Johns Hopkins University and a course on X-ray Imaging for UCLA's Extension Program. Currently Dr. Martz is writing a text book on Industrial X-ray Imaging.

Laura Parker

Department of Homeland Security

Laura Parker is a Program Manager in the Explosives Division of the Science and Technology Directorate at the Department of Homeland Security (DHS) as well as the Program Manager for the ALERT Center of Excellence, a DHS-sponsored consortium of universities performing research that address explosive threats lead by Northeastern University. She works on multiple projects for trace detection of explosives and algorithm development for improved explosives detection. Previous to her present position at DHS, Laura worked as a contractor providing technical and programmatic support of chemical and biological defense and explosives programs for several Department of Defense (DoD) offices. She also worked in several DoD Navy laboratories in the field of energetic materials. She obtained her Ph.D. in chemistry from the Pennsylvania State University.

Awareness and Localization of Explosives-Related Threats (ALERT)

A Department of Homeland Security Center of Excellence

Franco Rupcich

Engineering Contractor

Franco obtained his PhD in Biomedical Engineering from Marquette University in spring 2013 under the guidance of Taly Gilat-Schmidt, as well as several imaging researchers from the Center for Devices and Radiological Health (CDRH), FDA. His work focused primarily on investigating methods of reducing radiation dose to the breast during CT scans while maintaining diagnostic image quality as measured by objective, task-based metrics.

Previously, Franco has worked as a Systems Engineer at both Baxter Healthcare and Hospira performing requirements and risk management activities for dialysis machines and hospital infusion pumps.

Franco is currently a Lead CT Systems Engineer on the Image Quality and Dose team at GE Healthcare, where his primary duties include designing, developing, and testing new features for next generation CT scanner platforms.

Michael Silevitch

Northeastern University

Michael B. Silevitch is currently the Robert D. Black Professor of Engineering at Northeastern University in Boston, an elected fellow of the IEEE, and the Director of the Homeland Security Center of Excellence for Awareness and Localization of Explosives Related Threats (ALERT).

His training has encompassed both physics and electrical engineering disciplines. An author/co-author of over 65 journal papers, his research interests include laboratory and space plasma dynamics, nonlinear statistical mechanics, and K-12 science and mathematics curriculum implementation. Of particular interest is the study of the Aurora Borealis, one of nature's most artistic phenomena. Avocations include long distance hiking and the study of 17th Century clocks and watches.

Prof. Silevitch is also the Director of the Bernard M. Gordon Center for Subsurface Sensing and Imaging Systems (Gordon-CenSSIS), a graduated National Science Foundation Engineering Research Center (ERC). Established in September of 2000, the mission of Gordon-CenSSIS is to unify the methodology for finding hidden structures in diverse media such as the underground environment or within the human body.

Philip Top

Lawrence Livermore National Lab

Philip Top is a Research Engineer at Lawrence Livermore National Lab. He received his BS in 2002 and his Masters in 2004, before getting his PhD in 2007 from Purdue University in Signal Processing. His research areas include Remote Sensing classification and detection, Ultrawideband radar, Power grid modeling and simulation, and dynamic analysis.

Dong Hye Ye

Purdue University

Dr. Dong Hye Ye is a Postdoctoral Researcher in Electrical and Computer Engineering at the Purdue University. His research interests are in advancing machine learning for image processing. His publications have been awarded Best Paper at MICCAI-MedIA 2010 and Student Travel Grants at ISBI 2012, and PRNI 2012. During his PhD, Dong Hye conducted research at Section of Biomedical Image Analysis (SBIA) in Hospital of the University of Pennsylvania (HUP) and Microsoft Research Cambridge (MSRC). He received Bachelor's degree from Seoul National University in 2007 and Master's degree from Georgia Institute of Technology in 2008.

Jun Zhang

University of Wisconsin-Milwaukee

Jun Zhang received his Ph.D. in electrical engineering from Rensselaer Polytechnic Institute in 1985 and 1988, respectively. He joined the faculty of the Department of Electrical Engineering and Computer Science, University of Wisconsin-Milwaukee, and currently is a professor. His research interests include image processing and signal processing. He has been an associate editor of IEEE Trans. Image Processing and his research has received funding from NSF, ONR, State of Wisconsin, and industry.

11.5.1.4 Speaker Instructions

"Instructions for Researcher Presentations at the ATR Program Review and For Final Reports"

Instructions for Researcher Presentations at the ATR Program Review and For Final Reports

Version 4

Introduction

The purpose of this document is to present instructions for preparing your presentation for the program review (PR) and for your final report for the ATR project. The instructions are provided so that the presentations have uniform format and content in order to make it easier for the audience to understand your ATR.

Program Review Presentation Template

Please address the following topics in your presentation using the template that follows. Specifically, please follow the order provided and use no more than the specified number of slides.

Maximum # Slides	*Content*
1	Title slide - name of institution and researchers
1	Introduction to the institution and researchers
1	PD/PFA results – the following rows from pdpfa_log_pds.xls: 1 (header row), 16-33, 40, 42 (header row), and 43. Notes: 1. See section below entitled "Sample Format of PD/PFA Summary Slide" for sample format for this slide. 2. All results should be reported with the default values of precision and recall. See below for the defaults. The PD/PFA results should be for all ~200 scans. 3. You are only required to detect the targets (saline, clay and rubber) and the pseudo-target sheets. You are not required to detect pseudo target bulks (eg, the powders). Therefore, you can reduce PFA by not detecting pseudo target bulks at all.
1	Executive summary – High-level diagram and a couple of bullets describing your ATR and its motivation. Notes: 1. The audience may begin to ask questions at this point. Assume that you may have to describe your complete algorithm at this point.

ATR Project – Instructions for Program Review and Final Reports, Page 2

<table>
<tr><th>Maximum # Slides</th><th>Content</th></tr>
<tr><td>10</td><td>Technical description of the algorithm:
1. Philosophy and motivation for your approach
2. Segmentation including compensation for object splitting and merging
3. Features and their extraction
4. Correction for CT artifacts such as streaks and low-frequency shading
5. Classifier
6. How was shape used, if at all?
7. How were the following cases addressed differently: bulks, sheets?
8. Detection of targets in the presence of streaks and other CT artifacts
9. Detection of pseudo-target sheets
10. Relationship to the prior art
Notes:
1. Maximize use of figures and minimize use of equations. Recall that a picture (figure) is worth a 1000 words!
2. Show examples when your algorithm succeeds and fails at detecting objects
3. Do not present low-level details of your ATR; can you provide these details in reply to a question or after your presentation. You may include these details in backup slides.</td></tr>
<tr><td>5</td><td>ATR training:
1. How was over-training on supplied data prevented
2. How robust is the ATR to new types of targets
3. How were false alarms reduced?
Notes:
1. The slides discussing the algorithm itself and its training may be combined. However, the high-level description of the algorithm has to be presented first.
2. You may want to show 2D scatter plots when describing how your classifier is trained. The axes should be two features such as mass and density. You should show non-targets and targets in different colors.</td></tr>
<tr><td>1</td><td>Repeat Slide 3 with the PD/PFA results.</td></tr>
<tr><td>10</td><td>One slide for each of the ten cases described in the file entitled “presentation cases vxx.doc”. For each case:
1. Put case number and title of case in the title
2. Include the corresponding CT and ground truth (GT) images
3. Say if object was detected or not
4. If detected:
a. How the algorithm was designed to handle this case
b. Show label image for “slice”
5. If not detected
a. Why wasn’t it detected?
b. Show label image for “slice”
c. Show mass, recall, precision
d. What could be done to detect the target?
e. What would be the impact on PFA?
Notes:
1. May need more than one slide per case
2. The CT and GT images for these cases will be posted on the FTP site</td></tr>
<tr><td>2</td><td>Show at least two cases showing targets that were not detected; these can be contained in the 10 show cases or two other cases.</td></tr>
<tr><td>2</td><td>Show two cases where false alarms were created</td></tr>
</table>

ATR Project – Instructions for Program Review and Final Reports, Page 3

Maximum # Slides	*Content*
2	Discuss the strengths and weaknesses of the ATR including: 1. Limitations on types, densities, sizes, masses of targets, location, orientation, shape 2. What objects were missed and why are they missed? 3. Common false alarm objects and why were they 4. What is limiting further improvements in PD and PFA
1	How your ATR could be improved in the future. Topics that should be addressed include: 1. How PD and PFA can be improved 2. What specific cases need to be addressed 3. How more concentrations of saline and clay with different densities can be detected
1	Comments on the data (CT images, ground truth label images, scoring tools) and the process
1	What you learned by participating in this project.

Here are some guidelines for preparing your presentation.

1. All CT images should be displayed with a window width of 3000 MHU and a window level of 1000 MHU. The CT values of air and water on this scale are 0 MHU and 1000 MHU, respectively.
2. All images should be annotated with their SSN and slice number.
3. Check the agenda for the allocated time for your presentation. Prepare a presentation for one half of the time slot and assume that the rest of the time will be for discussion.
4. Bring your presentation on a USB memory stick in PowerPoint (PPT) or PDF. A PC will be provided with PowerPoint and Adobe Acrobat.
5. Do not present any SSI or classified information.
6. It is our intent to put your presentation into the public domain. You will be allowed to revise your presentation, and even redact information, after the program review.
7. Please let ALERT know if you have citations or reports for material that participants should read in advance of the symposium.
8. Contact Carl Crawford if you feel that the content or number of slides in the above table should be revised.
9. Do not present the background for the project. Carl will present that in his opening presentation.
10. You will be presenting to people who know how to develop ATRs for certified equipment. Do not teach the field to the audience.
11. A person should be able to read the presentation by him/herself.
12. Number all slides

Sample Format of PD/PFA Summary Slide

Target Type	Target Subtype or Form	Level of Difficulty	Num Targets	Num Detected	PD [%]
Target	All	All	407	386	94.8
Target	Clay	All	111	106	95.5
Target	Rubber	All	158	150	94.9
Target	Saline	All	138	130	94.2
Target	Bulk	All	270	259	95.9
Target	Sheet	All	137	127	92.7
Target	All	Low	77	74	96.1

ATR Project – Instructions for Program Review and Final Reports, Page 4

Target	Clay	Low	29	28	96.6
Target	Rubber	Low	22	21	95.5
Target	Saline	Low	26	25	96.2
Target	Bulk	Low	56	54	96.4
Target	Sheet	Low	21	20	95.2
Target	All	High	317	299	94.3
Target	Clay	High	82	78	95.1
Target	Rubber	High	125	118	94.4
Target	Saline	High	110	103	93.6
Target	Bulk	High	201	192	95.5
Target	Sheet	High	116	107	92.2
Num Non-targets	**Num FAs**	**PFA [%]**			
1371	147	10.7			

Default Values of Precision and Recall

Option	Default Argument	Synopsis
-p	0.5	Precision for bulk targets
-r	0.5	Recall for bulk targets
-q	0.2	Precision for sheet targets
-s	0.2	Recall for sheet targets
-t	0.5	Precision for bulk pseudo-targets
-v	0.5	Recall for bulk pseudo-targets
-u	0.1	Precision for sheet pseudo-targets
-w	0.1	Recall for sheet pseudo-targets

Final Report

Your final report should address the same points addressed in your presentation. That is, your report should mirror the slides/sections notes above your presentation. Your final report should also address the following points.

1. Biographies of team members
2. Bibliographies of relevant publications
3. Background
4. Discussion

Revision History

Ver.	Date	Who	Revisions
1	10/6	CRC	First draft

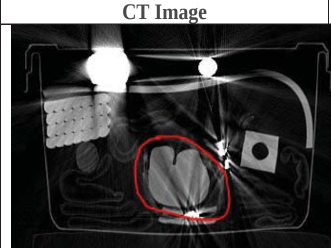
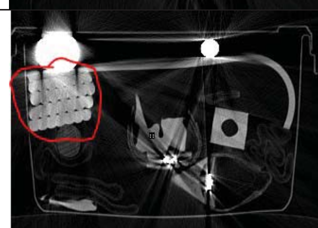
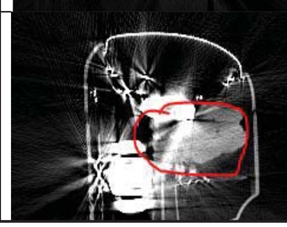
ATR Project – Instructions for Program Review and Final Reports, Page 5

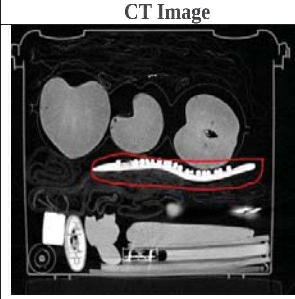
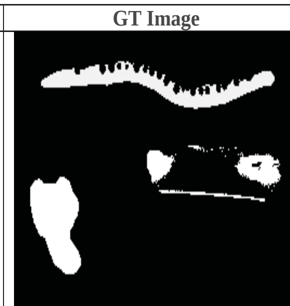
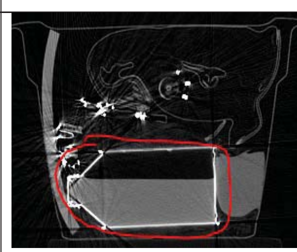
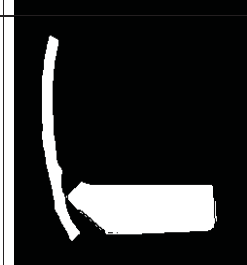
2	10/13	CRC	Based on comments from Karl, Rupcich and Birken; see tracking
3	10/15	CRC	Based on feedback from the ATR developers; see tracking
4	10/22	CRC	Based on review of first drafts of presentations; see tracking

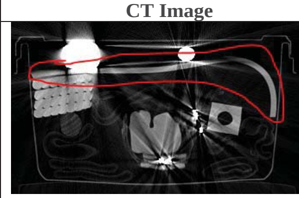
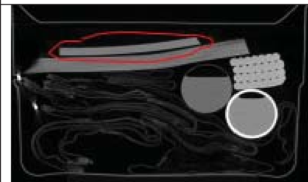
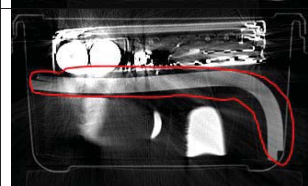
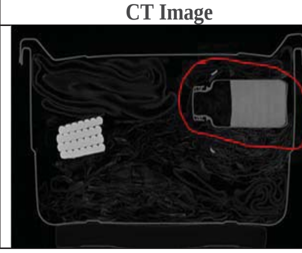
11.5.1.5 *Presentation Cases*

"Presentation Cases for Program Review"

Presentation Cases for Program Review

#	Case	Target	ID	Mass	SSN	Loc	Ort	Slice	CT Image	GT Image
1	Bulk with bad streaks caused by metal	Breast milk bag 10% Saline	6012	285	13	bbb	z	105		
2	Bulk with bad shading caused by beam hardening and scatter	clay	6051	286	13	abb	z	128		
3	Bulk inside electronics	clay	6150	290	35	aab	z	49		

#	Case	Target	ID	Mass	SSN	Loc	Ort	Slice	CT Image	GT Image
4	Bulk with texture	clay w/ glass beads	6193	410	193	bbc	x	198		
5	Bulk with density close to water (~5% saline)	5% saline – tin bottle	6163	274	63	baa	x	45		

#	Case	Target	ID	Mass	SSN	Loc	Ort	Slice	CT Image	GT Image
6	Sheet with bad streaks caused by metal, beam hardening and scatter	Rubber sheet 6.6 mm	6018	685	13	bcb	z	111		
7	Sheet laying on top of another flat object	3/8 rubber sheet on Elle magazine	6144	345	33	bca	x	46		
8	Object with lots of photon starvation	Merged rubber	7008	1360	11	bbb	z	94		
9	PT sheet based on thickness	Neoprene rubber sheet 3.2 mm	8026	350	18	bab	z	125		
#	Case	Target	ID	Mass	SSN	Loc	Ort	Slice	CT Image	GT Image
10	PT Powder (based on density, not mass)	TA_MH01 plastic bottle + powder	6026	277	12	cca	x	105		

Notation

1. # - case #
2. Case – purpose of example
3. Target – description of target
4. ID – target number
5. Mass – mass of target
6. SSN – scan ID
7. Loc – location code for target when packed
8. Ort – orientation code for target when packed
9. Slice – number of representative (image) shown in "Picture" column
10. CT image – CT slice for slice number show in "Slice" column
11. GT image – ground truth image for slice number show in "Slice" column
12. Notes – comments/notes about case

Notes:

1. Case 10: Not required to detect this object. But interesting to know what happens with this object
2. The snapshots of the CT and GT images shown in the above table can be found at: /eng_research_TO4/program-veview/cases

11.5.2 Presentations

11.5.2.1 Project Overview & CT Dataset (Crawford)

"ATR Project: Project Overview & CT Dataset"

ATR Project: Project Overview & CT Dataset

Carl Crawford, Csuptwo
David Castanon, Boston University
Clem Karl, Boston University
Harry Martz, Lawrence Livermore National Laboratory

1

So What? Who Cares?

- So What? … What was done …
  - Five ATRs developed: PD ~ 90%, PFA ~10%
    - Targets: saline, modeling clay, rubber sheets
    - Scanning: Medical CT; single-energy, 500 target scans
  - Automating scoring tools developed
    - Ground-truth labels - semi-automatically
    - Standardized reports
  - All of the above in public domain, by request
- Who cares?... To be determined by you if true …
  - Problem maps to security scanners
  - ATRs novel with respect to literature in public domain
  - Researchers available to contract to vendors
  - Students trained to work in industry
  - Third parties can work on unclassified, relevant projects
  - Scientific method continues to be applied for more improvements
  - TSA deploys better equipment derived from this project

2

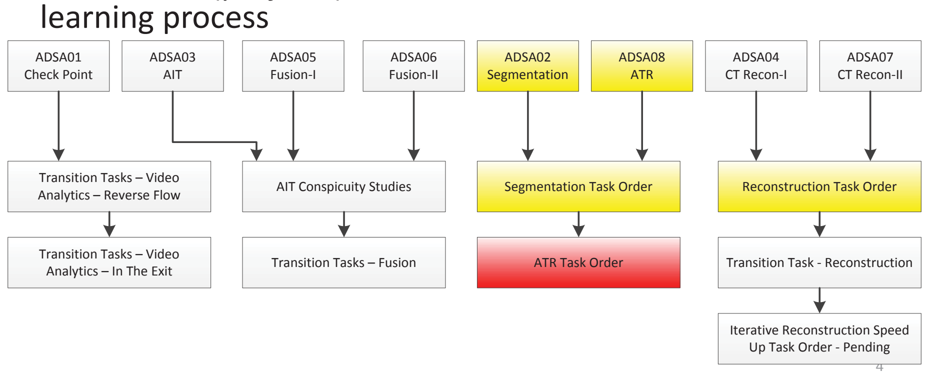
ADSA Workshop Format

- ADSA
  - Ask questions in real time
  - Interrupt speakers … "I do not understand"
  - Do not hold back – but play nice
  - Speakers expect this format
  - Time allocated in agenda for questions
  - Meeting will end at 5:30 PM
- No classified or SSI material
  - No mapping to security scanners
  - No saying something cannot be detected; say leads to high PFA
  - Minutes will be redacted if necessary

3

ADSA – Task Order Linkage

- DHS – involve third parties; augment vendors
- ADSA – develop methods to work with third parties
- Task Orders (projects) – learning process
- Start with CT-EDS segmentation
- Advance to reconstruction, ATR, AIT, video



4

Learned Post ADSA01

- Problems
  - TSA requirements and data from vendor equipment cannot be provided to 3rd parties; 3rd parties cannot test at TSL
    - Detection requirements are classified
    - Data from deployed equipment are SSI or classified, and are under export control
    - DHS/TSA policies do not allow TSL to test components (e.g., an ATR) separate from a complete scanner
    - Privacy concerns with scans on AIT equipment.
    - Business interests of the vendors should be protected
  - There was no publicly available set of images that are representative of challenging ATR problems for explosive detection systems.
- Solutions
  - For EDS, detecting Coke in presence of Pepsi equivalent to detecting explosives in presence of peanut butter
  - Scan targets (non-threats) on non-security scanner; create project-specific detection requirements

Working with 3rd parties much more difficult than it looks

5

ATR Project - Bottom Line

- Develop automated target recognition (ATR) algorithm to detect targets in scans on a medical CT scanner
- Input = 3D CT data + projections
- Output = 3D label image (i.e., a “mask”) indicating pixels of detected targets
- Targets
  - Saline, modeling clay, rubber sheets
- Detection of objects in ATRs
  - Defined in terms of recall and precision
  - Determined using a ground truth label image and automated scoring tools
- Goals for ATRs
  - Probability of detection (PD) > 90% (all targets + pseudo target sheets)
  - Probability of false alarm (PFA) < 10%

6

ATR Project - Bottom Line (II)

- Prior art methods are proprietary and classified
  - Will not know if results are better
- Success
  - Understanding ATR problem
  - Being able to work with vendors in the future
  - Solving detecting difficult cases more important than trying to detect easy cases
  - Putting database into public domain
- Sample ATR, scoring programs supplied to reduce development efforts
- Bibliography supplied of prior art

7

Project Team

- ATR developers
  - Purdue University: Dong Hye Ye, Charlie Bouman, Pengchong Jin
  - Massachusetts General Hospital/Harvard; Synho Do
  - University of Tennessee: Jens Gregor
  - University of Wisconsin: Jun Zhang, Laura Drake, Hongquan Zuo
  - Lawrence Livermore National Laboratory: Philip Top, Ana Paula Sales, Hyojin Kim, Timo Bremer, Steve Azevedo, Harry Martz
- Data collection, database
  - Doug Boyd, Sam Song, Tip Partridge, Telesecurity Sciences
  - Rick Moore, Alyssa White, Massachusetts General Hospital/ALERT
  - Steve Skrzypkowiak, DHS
- Tools, ground truth, database
  - Franco Rupcich, Independent consultant (PHD with Taly Gilat-Schmidt)
- Technical leadership
  - David Castanon, Clem Karl, Boston University
  - Carl Crawford, Csuptwo
  - Harry Martz, Lawrence Livermore National Laboratory
- Programmatic leadership
  - Michael Silevitch, John Beaty, Ralf Birken, Northeastern University/ALERT

8

Targets

- Saline: ~1035 – 1150 MHU
  - 3.5%, 10%, 15% concentrations
  - Container not part of target; only the saline
- Modeling (polymer) clay
- Rubber : ¼" thickness (minimum) sheets + other rubber in bags (bulks)
- Minimum mass: 250 g (physical, not CT)
- Maximum mass: none
- Form: contiguous

9

Targets

10

Pseudo Targets

- Used to test lower mass and density thresholds in future
- Saline, clay, rubber sheets
  - 125 g < mass < 250 g
- Powders:
  - Density < 1 g/cc
  - Mass > 125 g
- Sheets
  - < ¼” thick
- Pseudo targets do not count in PD or PFA calculations
  - Exception: Required to detect pseudo target sheets

11

Object Philosophy Issue

- Scanned stacks of N rubber sheets
  - Is this one object or N separate objects?
- Some touching targets merged into one target
  - Subjective
  - Leads to ceiling/floor on PD/PFA
  - Penalize ATR if separated

12

Non-Targets

- Stream of commerce items
  - Food
  - Drinks
  - Electronics
  - Magazines
  - Water

13

Packing

- Targets packed with different
  - Shapes
    - Most bulks: saline and clay
    - Most sheets: rubber
  - Containers – metal, plastic, glass
  - Concealment
  - Clutter
  - Location
  - Orientation
  - Texture
- Plastic bins

14

Packing Examples

15

Level of Difficulty Per Target

- **High LOD** if
  - Target is touching a non-target with similar density
  - A *discontinuous* shape was created because of object philosophy. This subjective term means that the target could have been assembled from two more pieces of the base material (saline, rubber and clay).
  - Streaks pass through the target
  - Low frequency shading (caused by beam hardening and scatter) depress the CT values by more than 10%
  - Sausage-like scanned with its long axis contained in the x-y plane.
  - Sheet targets scanned with most of its mass contained in the x-y plane.
  - Sheets are rolled up
  - Texture is present in the target. For example, clay mixed with glass beads.
  - High electronic or quantum noise present in the CT images
  - Pseudo-target
  - Saline with concentration of less than 3.5% salt

Concentrate on High LOD

16

Scanning

- Scan on Imatron medical CT scanner
  - Same scanner and protocol as reconstruction project
  - Single energy
  - Raw data collected
  - Reconstructed with offline version of vendor's reconstruction algorithm (xrec)
- Scans: 188 of bins
  - 421 targets
  - 75 pseudo targets
  - 1371 non-targets

17

Database

- CT scans (~50k images)
- Raw data
- Ground truth was created only for targets and pseudo targets, not for non-targets
- Information on packing
  - List of objects and characteristics (mass, size, etc.)
  - Pictures – targets, non-targets
  - Videos of bins when unpacked
  - Objects, bins have IDs

18

ATR Definition

- Input to ATR: 3D CT images + projections
  - Objects are not segmented
  - Feature are not extracted
  - Raw data cannot be reconstructed
  - Ground truth only to learn about targets
- Possible Functions: segmentation, feature extraction, CT correction, classification
  - Not required – could find targets in projections
- Output from ATR: label images, log file

19

ATR Output: Label Images

- Label Image: *An image generated either by an ATR or by a ground truth generating program indicating to which **label** a pixel belongs*
- A label image pixel can be assigned to only one object
- Values:
  - 0 = background
  - Label/tag for object for ground truth
  - >0 = object number from ATR

20

Ground Truth

- Label images showing locations of targets
- Used for
  - Target analysis (feature extraction) by researchers
  - Automated scoring
- More in Franco's talk

21

ATR Implementation

- Sample ATR supplied based on Karimi's code for Segmentation Project
  - Demonstrate reading FITS images, writing label images and log files
  - Can rip out algorithm and insert your own
  - C/Linux
- Do not over-train on the data or use illegal features
- Execution time out of scope

22

ATR Overview (Prior Art)

Multi-Path Segmentation → Feature Extraction → CT Correction → Classifier

Multi-Path Segmentation:
Sheet filter/path
Bulk filter/path
Weapons filter/path

Feature Extraction:
Mass
Density
Z-effective
Texture
Volume

CT Correction:
Orientation
Resolution

Classifier:
Non-linear thresholds

Do something other than prior art

23

Features (Prior Art)

- Mass
- Mean: LAC, Zeff
- Standard deviation: LAC, Zeff
- Histograms
- Higher-order moments
  - Skew, kurtosis, entropy
- Texture
  - Wavelets

Your responsibility to determine relevant features

24

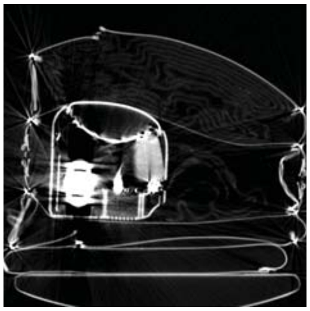
Illegal Features

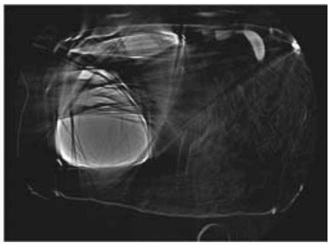
- Shape (except for minimum sheet thickness)
  - Can be used for separate paths
- Orientation
- Location
- Maximum mass, volume
- Container type

Will review your features during course of the project

25

Typical Image Quality





- Artifact types
  - Shading
  - Streaks
  - Noise
  - Blurring
  - Rings
- Artifacts lead to
  - Merging of objects
  - Splitting of objects
  - Imprecise density, volume, mass, shape

How are IQ issues handled in ATR?

26

ATR Testing

- Run by performer at their site
  – Too difficult to execute at ALERT
  – Honor system not to over-train
- Results scored by provided tools
  – PD/PFA
  – Log files to report results

27

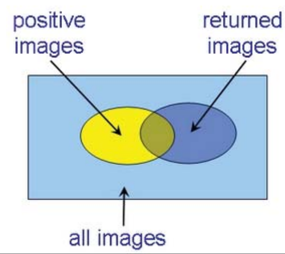
Performance Metrics

- Goal: PD > 90%, PFA < 10%
  – All targets
  – Pseudo target sheets
  – Concentrate on high LOD cases
- PD = # targets detected / # targets scanned
- PFA = # false alarm objects / # non-targets scanned
- Detection based on recall/precision

28

Detection and False Alarm

- Detection is when ATR label overlaps ground truth label
  – Overlaps at least 50% of ground truth (recall, r)
  – At least 50% of label overlaps ground truth (precision, p)
- Target sheets: p=r=0.2; Pseudo target sheets: p=r=0.1



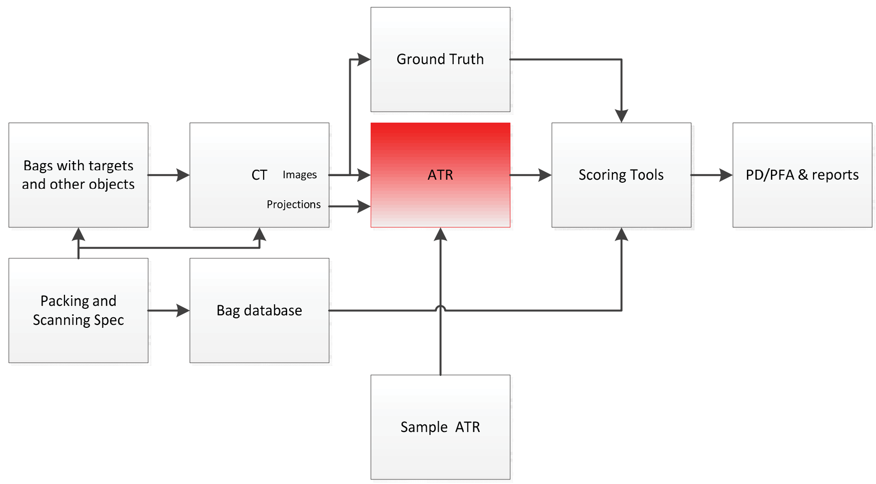
29

Support Functions (Tools)

- Franco Rupcich coding in C
- Sample ATR
  - Reading image
  - Writing results (label, log files)
  - Revised as necessary
  - Replace ATR functions with your own
- Scoring software
  - Detection using recall/precision
  - PD/PFA
- Simulated images for validation
- Image conversion to FITS
- Ground truth - Mevislab

30

Block Diagram



31

Program Success

- Problem maps to security scanners
- ATRs novel with respect to literature in public domain
- Researchers available to contract to vendors
- Students trained to work in industry
- Third parties can work on unclassified, relevant projects
- Scientific method continues to be applied for more improvements
- TSA deploys better equipment derived from this project

32

Uniqueness

- Bibliography and copies prior art provided
  - Patents, presentations, reports, articles
  - Vendors may be doing something else
- Do not replicate prior art
- Extend sample ATR

33

Collaboration

- Researchers encouraged to collaborate with other team members
  - Tools, image formats
  - Not on algorithms
- Segmentation expertise from segmentation project

34

Out of Scope

- Execution speed + computational expense
- Equating to performance of equipment deployed by TSA
- Reconstructing CT projection data; do not want to re-run the Reconstruction Project (Task Order 3).

35

Publications/Patents

- Publications permitted and encouraged
- Prior review required by ALERT
  – Process denoted REAP
  – Use “leads to higher false alarms”
  – Do not say “cannot detect”
- May obtain patents on work

36

Furnished to ATR Developers

- Specifications
  - Technical top-level
  - Tools – scoring, sample ATR
  - Scanning/packing
- Database
  - CT images
  - Ground truth labels
  - Sample ATR
- Code
  - Automatic scoring
  - Sample ATR
- Access to subject matter experts

37

Performer Future

- Developing vendor-neutral ATRs
- Working with vendors
- Work on other modalities
- Prediction of detection capability of future scanners

38

Data share access

- Data Resource warehoused on networked BU server
  - Raw data, reconstruction software, documentation, scanner models, simulations
- Process for obtaining data
  - Request NDA from ALERT
  - Obtain account to access network data share site
  - Must agree to have publications reviewed ("REAP'ed") for problematic wording
- Access to data is available after project ends
- Researcher results & documentation can be posted as well

39

Differences from TSL Certification

- Saline, clay, rubber, not explosives
- Medical CT scanner
- Bins, not bags
- PD/PFA per object type, not per bag
- Comingled PD/PFA runs
  - Missed detection can lead to false alarm because of recall/precision
- Automated scoring with recall/precision
- Testing open, not blind
- Test/train on same set
- Statistical relevance not an issue

40

Feedback

- Vendor panel

David Lieblich	Analogic
Kam Lin Wong	Reveal
David Perticone	L-3 Communications
James Connelly	IDSS
Carl Bosch	Surescan
Ling Tang	Rapiscan
Piero Landolfi	Morpho Detection
Christopher Gregory	Smith Detection

- Questionnaire for all participants
  - www.surveymonkey.com/s/TaskOrder4
- Specifications, results online
  - https://myfiles.neu.edu/groups/ALERT/strategic_studies/TaskOrder04/

41

Acknowledgements

- Laura Parker, DHS, funding
- Vendor feedback
  - In particular, Matthew Merzbacher, Morpho Detection
- Jeffrey Kallman, Steve Azevedo, LLNL, technical support
- Heartscan, South San Francisco, Imatron scanner
- ALERT staff
  - Melanie Smith(**), Teri Incampo, Kristin Hicks, Deanna Beirne, Anne Magrath, Sara Baier

42

Super Special Acknowledgements

- The ATR initiative would not have been a success without the research groups. The success of this project is due 99.99% to their contributions.
- We extend our heartfelt thanks to them for their participation and working the project team to fix issues when they became evident.

43

Indemnification

- All problems, issues and bugs are the responsibility of Carl Crawford, David Castanon, Clem Karl and Harry Martz, not the researchers and support staff
- ATR presentations have uniform format and content in order to make it easier for the audience to understand the ATRs
  - Format may have bugs – TBD
  - Be patient

44

So What? Who Cares?

- So What? … What was done …
  - Five ATRs developed: PD ~ 90%, PFA ~10%
    - Targets: saline, modeling clay, rubber sheets
    - Scanning: Medical CT; single-energy, 500 target scans
  - Automating scoring tools developed
    - Ground-truth labels - semi-automatically
    - Standardized reports
  - All of the above in public domain, by request
- Who cares?... To be determined by you if true …
  - Problem maps to security scanners
  - ATRs novel with respect to literature in public domain
  - Researchers available to contract to vendors
  - Students trained to work in industry
  - Third parties can work on unclassified, relevant projects
  - Scientific method continues to be applied for more improvements
  - TSA deploys better equipment derived from this project

45

BACKUP SLIDES

46

DHS Goals

- Vendors doing an excellent job
- But, need
  - Increase probability of detection (PD)
  - Decreased probability of false alarm (PFA)
  - Detect more threats including wide-variation of home-made explosives (HMEs)
  - Reduced mass
  - Reduced total cost of ownership

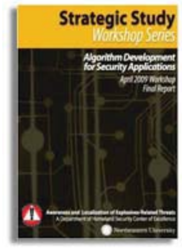
47

DHS Tactics

- Augment abilities of vendors with 3rd parties
  - Academia
  - National labs
  - Industry other than the vendors
- Create centers of excellence (COE) at universities
  - ALERT, Northeastern University
- Hold workshops to educate 3rd parties and discuss issues with involvement of 3rd parties
  - Algorithm Development for Security Applications (ADSA)
  - 11 ADSAs to date

48

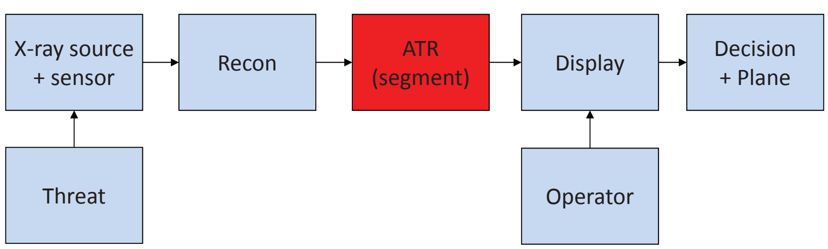
ADSA - Recommendations



- ADSA01 Recommendations
  - Execute research projects (grant challenges, task orders, initiatives)
  - Segmentation first – easiest task
    - Begin to learn how to work with 3rd parties
  - Reconstruction second
    - Difficult to get projection data and parameters
    - Difficult to assess results
  - ATR third
- Recommendations refined at subsequent ADSAs and with task orders

49

EDS Diagram



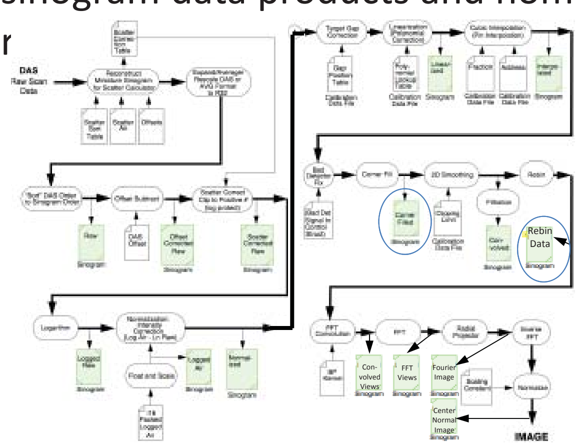
50

- Questions for panel
- Questions for participants

51

Imatron C300 Data and Software

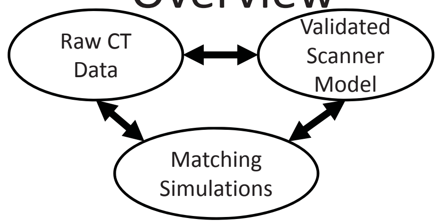
- Access to software processing chain, raw sinogram data products and nominal r



Examples:

- Filled fanbeam
- Rebinned parallel
- Corresponding projection models

The TO3 /TO4 Data Resource Overview



- The only open access X-ray security resource for third parties
- Based on Imatron C300 medical scanner
- Mixed mono and dual energy
- Scanned data of "security interest" (i.e. not medical)
- Validated scanner model
  - U. Chicago: Patrick La Riviere, Phillip Vargas
- Coupled validated simulation
  - Marquette University: Taly Gilat-Schmidt

53

Data Resource Goals

- Raw data, images, models, tools, documentation in public domain
  - Allow third parties to develop advanced algorithms
- All data on FTP site
- Available to all researchers after signing agreement with ALERT
  - Cannot redistribute data
  - Publications/presentations reviewed

54

More Performance Criteria

- Minimize use of special cases (corner cases)
- Feature space chopped up
  - Over-training
- Extensible for new targets

55

11.5.2.2 Tools and ground truth (Rupcich)

"ATR Project (Task Order 4) Program Review: Database, Ground Truth, & Software Tools"

ATR Project (Task Order 4)
Program Review
Database, Ground Truth, & Software Tools

Franco Rupcich, PhD

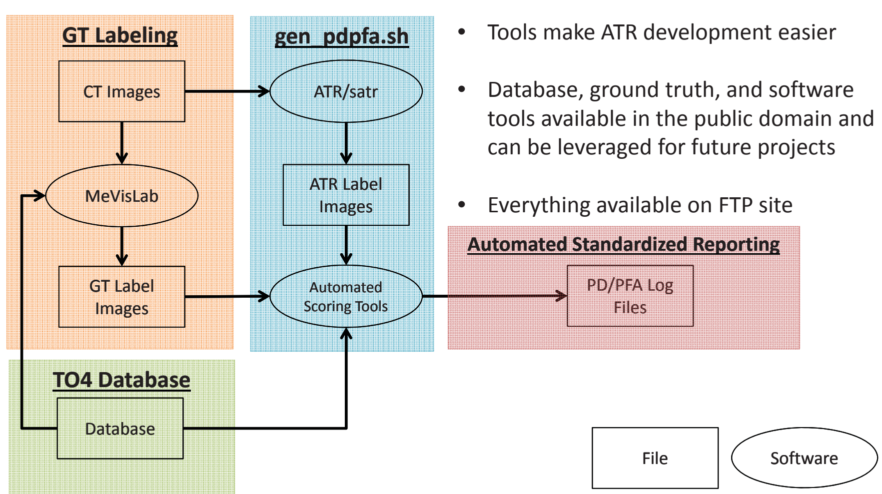
6 November 2014

Conclusions

So What?

Who Cares?

- Tools make ATR development easier
- Database, ground truth, and software tools available in the public domain and can be leveraged for future projects
- Everything available on FTP site



10/21/14 SW Tools & Ground Truth 2

Biography

- Education (Marquette University, Milwaukee, WI)
  - PhD, Biomedical Engineering
- Research
  - Radiation dose reduction in medical CT
  - Objective image quality metrics
  - Energy-resolved CT
- Career
  - Lead CT Systems Engineer at GE Healthcare (Waukesha, WI)

10/21/14 SW Tools & Ground Truth 3

TO4 Database

- Excel workbook describing
  - Objects (mass, material, target type, etc.)
  - Packing/Scanning (objects in each SSN, location, orientation, etc.)
- Includes packing videos and pictures of targets
- Available on FTP site

10/21/14 SW Tools & Ground Truth 4

Object Database

- Worksheet in Excel containing info for each object
  - Object IDs
  - Type
  - Sub-type
  - Form
  - Dimensions
  - Mass

	A	B	D	E	F	K	L	M	P
1	ID#	Object	typ	sub-ty	form	dim-1 (mn	dim-2 (mn	dim-3 (mr	m (g)
238	6038.0	plastic bottle 125 ml	n	e	bulk				20
239	6039.0	plastic bottle 125 ml	n	e	bulk				20
240	6040.0	%5 saline in 6039	pt	s	bulk				138
241	6041.0	clay cut from 6034	t	c	bulk	65	40	150	549
242	6042.0	10% saline in 6038	pt	s	bulk				137

2/5/2015 TO4 Monthly Status Update 5

Packing Database

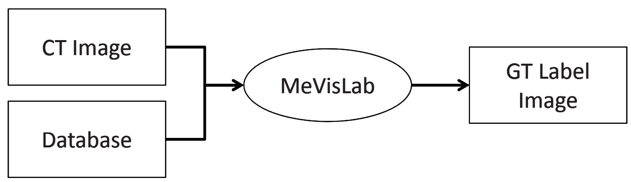
- Worksheet in Excel containing info for each object in each scan
  - Object IDs
  - Boundary box of each target/pseudo-target
  - Level of Difficulty of each target/pseudo-target
  - Object database info for each object

	A	B	C	E	F	G
1	SSN	Object I	Target Desc. (R, S, C, P, N)	Location	Orientation	LOD
2	4	6003	bin			
3	4	6004	Rubber mallet	bbc	z+	
4	4	6002	Saline	bca	y+	High
5	4	6006		bab	x+z+	
6	4	6010		cac	x+z+	

2/5/2015 TO4 Monthly Status Update 6

GT Labeling Overview

- Purpose
  - GT label images required as reference against which performance of an ATR can be scored
- Methods
  - Each GT label image was created via semi-automated segmentation and masking of each target in each CT image
- All GT label images and source code available on FTP site

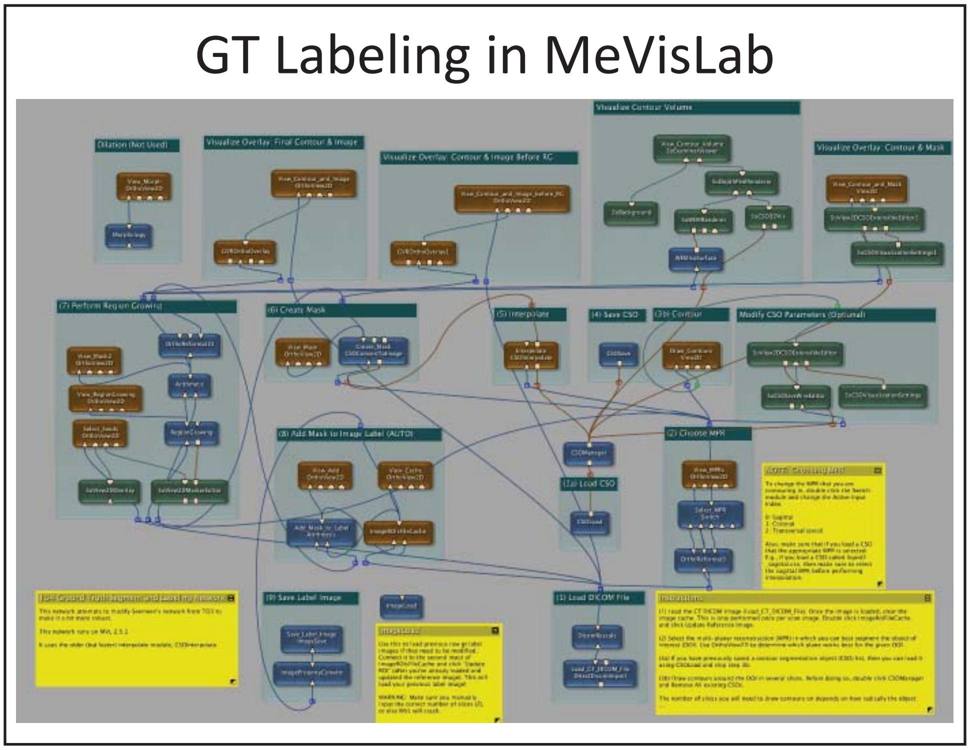


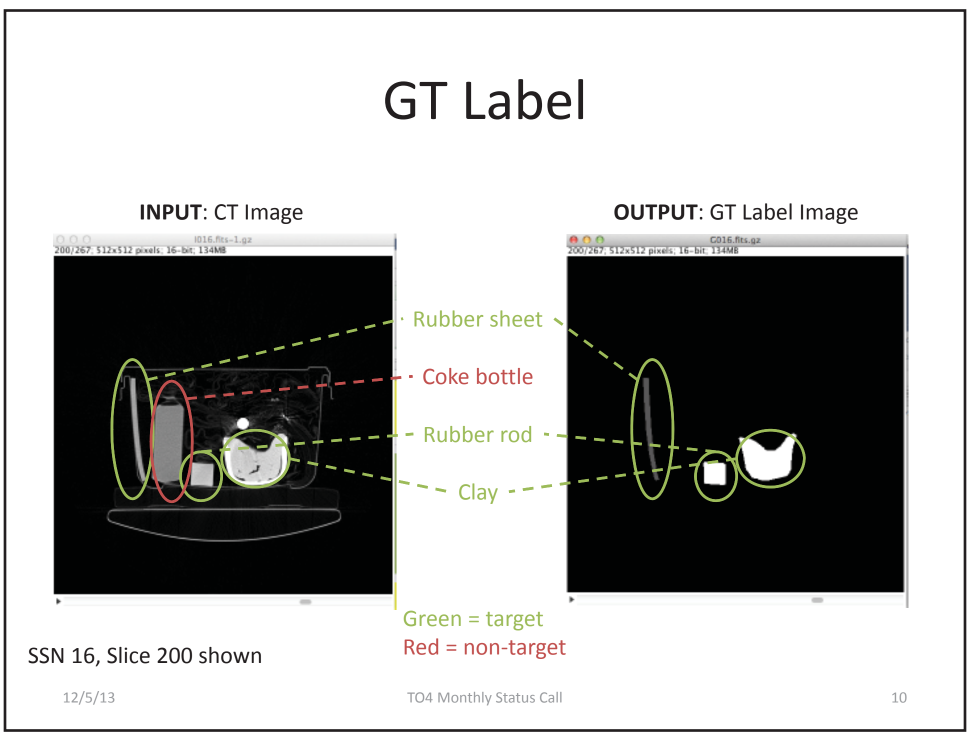
10/21/14 SW Tools & Ground Truth 7

GT Labeling Methods

- MeVisLab
- Semi-automated labeling program already developed by Seemeen Karimi for TO1 and modified for TO4
- Image processing modules readily available
  - Semi-automated segmentation
  - **Interpolation of contours across slices**
  - Region growing
  - Masking

12/5/13 TO4 Monthly Status Call 8


GT Labeling in MeVisLab


GT Label
INPUT: CT Image
OUTPUT: GT Label Image
Rubber sheet
Coke bottle
Rubber rod
Clay
Green = target
Red = non-target
SSN 16, Slice 200 shown
12/5/13
TO4 Monthly Status Call
10

GT Labeling Issues

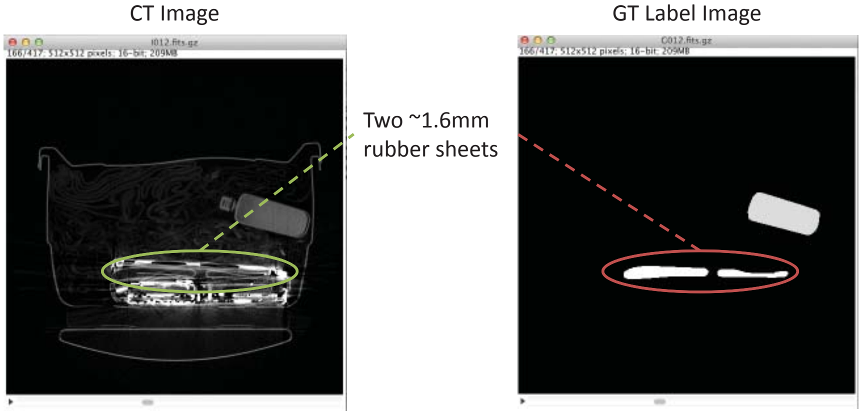
- MeVisLab often crashes during interpolation step
  - SOLUTION: draw more contours for targets that change radically between slices
- Thin sheets are difficult to segment
  - SOLUTION: reduced P/R specs for thin PT sheets
- CT artifacts make segmenting/region growing difficult
  - SOLUTION: segment target in many slices
- Time-consuming
  - ~20-40 minutes per bag

12/5/13 TO4 Monthly Status Call 11

GT Labeling Issues: Thin Sheets

CT Image

GT Label Image



SSN 12, Slice 166 shown

12/5/13 TO4 Monthly Status Call 12

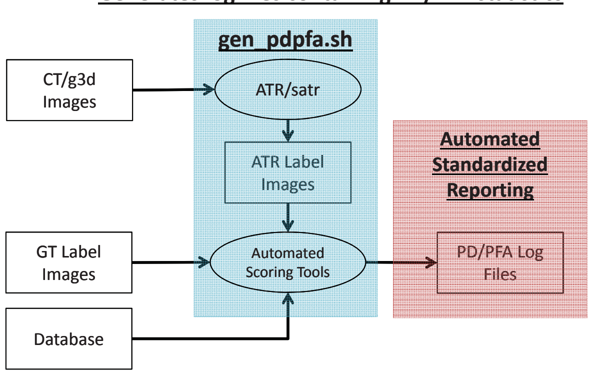
Sample ATR (satr)

- Performs erosion and CCL on CT image
- Calculates the mass of objects and has a classifier based only on mass and density
- Outputs label image and log file
- Served as basis for ATR code (file I/O, log files, etc.)
- Notes
  - Code originally written by Seemeen Karimi and modified by Carl Crawford and Franco Rupcich for TO4
  - Can be used by researchers as basis of their ATR

12/5/13 TO4 Monthly Status Call 13

gen_pdpfa.sh

1. **Runs an ATR/satr** on a set of CT images to produce label images
1. **Scores the label images** against GT
1. **Generates log files containing PD/PFA statistics**



- Software code
  - gen_pdpfa.sh is Bash shell script
  - Automated scoring tools in C
  - runs on Mac OS X and 32-/64-bit Linux
  - available on FTP site
- Researchers ran code at their respective sites

10/21/14 SW Tools & Ground Truth 14

Standardized Reporting

- pdpfa pds log file [.xls]
  - PD/PFA statistics

Log file used by researchers to report PD/PFA

- pdpfa summary log file [.txt]
  - summary PD/PFA information
- pdpfa detections log file [.xls]
  - indicates whether each target was detected or missed
- pdpfa false alarm log file [.xls]
  - information for each false alarm produced

10/21/14 SW Tools & Ground Truth 15

Standardized Reporting: pdpfa_log_pds.xls

PD info based on

- Type
- Subtype
- Form
- LOD

PFA info

- PFA
- Average number false alarms

	A	B	C	D	E	F
1	Target Type	Target Subtype or Form	Level of Difficulty	Num Targets	Num Detected	PD [%]
2	All	All	All	480	337	70.2
3	All	Clay	All	121	96	79.3
4	All	Rubber	All	168	143	85.1
5	All	Saline	All	157	97	61.8
6	All	Powder	All	34	1	2.9
7	All	Bulk	All	333	212	63.7
8	All	Sheet	All	147	125	85
9	All	All	High	390	257	65.9
10	All	Clay	High	92	72	78.3
11	All	Rubber	High	135	110	81.5
12	All	Saline	High	129	74	57.4
13	All	Powder	High	34	1	2.9
14	All	Bulk	High	264	153	58
15	All	Sheet	High	126	104	82.5
16	Target	All	All	407	312	76.7
17	Target	Clay	All	111	87	78.4
18	Target	Rubber	All	158	140	88.6
19	Target	Saline	All	138	85	61.6
20	Target	Bulk	All	270	190	70.4
21	Target	Sheet	All	137	122	89.1
22	Target	All	Low	77	67	87
23	Target	Clay	Low	29	24	82.8
24	Target	Rubber	Low	22	22	100
25	Target	Saline	Low	26	21	80.8
26	Target	Bulk	Low	56	46	82.1
27	Target	Sheet	Low	21	21	100
28	Target	All	High	317	232	73.2
29	Target	Clay	High	82	63	76.8
30	Target	Rubber	High	125	107	85.6
31	Target	Saline	High	110	62	56.4
32	Target	Bulk	High	201	131	65.2
33	Target	Sheet	High	116	101	87.1
34	Pseudo-target	All	High	73	25	34.2
35	Pseudo-target	Clay	High	10	9	90
36	Pseudo-target	Rubber	High	10	3	30
37	Pseudo-target	Saline	High	19	12	63.2
38	Pseudo-target	Powder	High	34	1	2.9
39	Pseudo-target	Bulk	High	63	22	34.9
40	Pseudo-target	Sheet	High	10	3	30
41						
42	Num Non-targets	Num FAs	PFA [%]			
43	1371	330	24.1			
44						
45	Num FAs	Num Scans with FAs	Avg Num FAs			
46	330	140	2.36			

2/5/2015 TO4 Monthly Statu

Other Work

- Specifications documents (available on FTP site)
  - Top-Level Spec - high-level specs for TO4 project
  - SW Tools spec - detailed specs for each SW tool
  - Level-of-difficulty – specs for LOD classifications (Low, High)
  - Ground Truth Labeling – spec/instructions for GT labeling program
- LOD Assignments
  - Worked with Stephen Skrzypkowiak to verify all LOD assignments
- gtver
  - Program to verify integrity of GT labels against database
- g3d test images
  - Two test images and corresponding GT label images are available
  - More can be created upon request

Lessons Learned

- Scanning and Database creation
  - Record as much information during scan-time as possible
  - Revise data collection procedure based on “Lessons Learned” from collecting and using sample data first
- Precisely defining requirements up-front saves time later
  - Changing GT labels
  - Changing code
  - Object philosophy
- GT labeling
  - Could be small project in and of itself
  - Future consideration: create GUI using ITK/VTK C++ tools with image processing modules catered specifically toward future TO projects...?

Future

- Database, GT labeling program, and automated scoring tools may be useful for future DHS and TSA projects
  - AT2
  - AIT
  - Cargo applications

10/21/14 SW Tools & Ground Truth 19

Questions?

10/21/14 SW Tools & Ground Truth 20

Backup slides

10/21/14 SW Tools & Ground Truth 21

Sheets

Z < ¼" (6mm): sheet pt

¼" (6mm) < z <= 3/8" (10mm): sheet t

Z > 3/8" (10mm): bulk t

10/21/14 SW Tools & Ground Truth 22

Label Images

- 3D image that indicates if a pixel in a CT image corresponds to a target

- Output from
  - Ground Truth generating program (Ground Truth Label Image
  - ATR program (Label Image)

10/21/14 SW Tools & Ground Truth 23

Precision and Recall

<u>Precision</u>

The fraction of a label declared by an ATR that overlaps with a target as declared by in the ground truth label image

<u>Recall</u>

The fraction of a target, as declared in the ground truth label image, that overlaps with a label detected by an ATR

10/21/14 SW Tools & Ground Truth 24

Detection, Miss, and False Alarm

- A detection occurs when an alarm declared by an ATR matches the ground-truth for a target. The term match is defined in terms of recall, R, and precision, P
- A false alarm occurs when an ATR creates a label that does not match the requirements for a detection
- A miss occurs when an ATR produces no label that satisfies the precision and recall specifications for a target in the ground truth label image

06/05/14 TO4 Monthly Status Update 25

Incomplete Detections

- A detection that occurs for the values of precision and recall multiplied by the factor, *alpha*
- The default value of alpha is 0.0.
  - This implies that an ATR label that intersects a GT label by at least one pixel meets the requirement for an incomplete detection.
  - If an ATR label meets the requirements for a detection for a GT label, then it will not be counted as an incomplete detection for that GT label
- Note the following:
  - Incomplete detections ***do not*** count as a detection
  - Incomplete detections ***do*** count as false alarms
- ATR labels that generate incomplete detections can be found under the “False Alarms” section of the *dder* Summary Log File, as well as in the *dder* False Alarms Log File.
  - Any false alarms that have “intersecting GT labels” reported are incomplete detections

06/05/14 TO4 Monthly Status Update 26

Incomplete Detections

- Implemented to help identify ATR labels that overlap GT labels but do not fully meet precision/recall specs for a detection
- May lead to clues to help increase PD or decrease PFA

06/05/14 TO4 Monthly Status Update 27

TO4 Database

Excel workbook containing information about all objects and SSNs for the TO4 project

/eng_research_TO4/database/**to4 database v30.xlsx**

- **Object Database** -- "objects" sheet in the Excel file
- **Packing Database** -- "Packing" sheet in the Excel file
- **Height Database** -- "Scan info" sheet in the Excel file
- **SSN Filtering** -- "SSN Filter" sheet in the Excel file
  - Allows filtering of SSNs based on contents
    - Type
    - Sub-type
    - Form
    - LOD

NOTE
satr, dder, and pdpfa use .csv versions of these databases, which are version controlled and bundled with each release of the tools

NOTE
The height database is used only by satr to zero out the patient table

2/5/2015 TO4 Monthly Status Update 28

Height Database

Contains height of patient table for each SSN

Used by satr to zero out patient table in CT image to avoid false alarms

	A	B
1	SSN	height
2	4	383
3	5	383
4	6	383
5	7	383
6	8	407
7	9	407
8	10	384
9	11	384
10	12	413
11	13	384
12	14	384
13	15	454
14	16	384
15	17	427
16	18	391
17	19	382
18	20	404
19	21	404
20	22	389
21	23	383
22	24	413
23	25	385
24	26	383

10/21/14 SW Tools & Ground Truth 29

TO4 Database: SSN Filtering

Find SSNs that contain specified
- # targets/pseudo-targets
- # targets/pseudo-targets of certain sub-type
- # targets/pseudo-targets of certain form
- # targets/pseudo-targets of certain LOD
- Any combination of the above

	A	V	W	X	Y
1		Level of Difficulty			
2		Low	High		
3	SSN	# Targets	# Targets	# Pseudo-targets	# Targets & Pseudo-targets
4	4	0	1	0	1
5	5	0	1	0	1
6	6	0	2	0	2
7	7	0	4	0	4
8	8	0	1	1	2

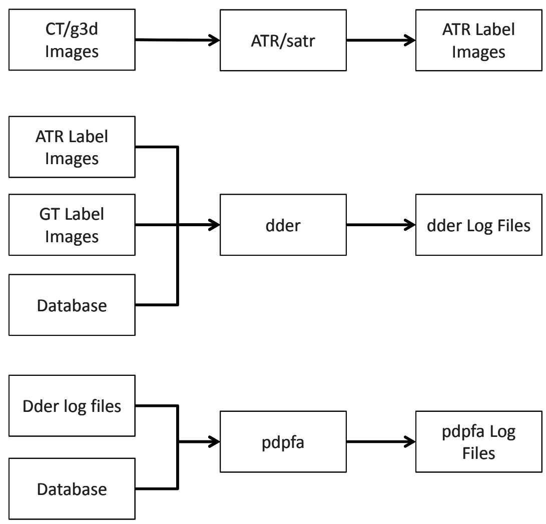
Can be used to help create desired subsets of SSNs, for example:
- SSNs with > 0 pseudo-targets
- SSNs with > 0 targets/pt with high LOD
- SSNs with > 0 rubber sheet targets with high LOD

	A	P	Q	R	S	T	U
1		Form					
2		Bulk			Sheet		
3	SSN	# Targets	# Pseudo-targets	# Targets & Pseudo-targets	# Targets	# Pseudo-targets	# Targets & Pseudo-targets
4	4	1	0	1	0	0	0
5	5	1	0	1	0	0	0
6	6	1	0	1	1	0	1
7	7	3	0	3	1	0	1
8	8	1	1	2	0	0	0

	A	B	C	D	E
1		Type			
2					
3	SSN	# Objs	# Targets	# Pseudo-targets	# Target & Pseudo-targ
4	4	5	1	0	1
5	5	5	1	0	1
6	6	4	2	0	2
7	7	10	4	0	4
8	8	13	1	1	2

	A	F	G	H	I	J	K	L	M	N	O
1		Sub-type									
2		Rubber			Saline			Clay			Powder
3	SSN	# Targets	# Pseudo-target	# Targets & Pseudo-targets	# Targets	# Pseudo-target	# Targets & Pseudo-targets	# Targets	# Pseudo-target	# Targets & Pseudo-targets	# Pseudo-target
4	4	0	0	0	1	0	1	0	0	0	0
5	5	0	0	0	1	0	1	0	0	0	0
6	6	1	0	1	1	0	1	0	0	0	0
7	7	1	0	1	3	0	3	0	0	0	0
8	8	0	0	0	1	0	1	0	0	0	1

Block Diagrams: ATR/satr, dder, pdpfa



10/21/14 SW Tools & Ground Truth 31

Log File: satr_log.txt

Info about labels produced by SATR

- Label IDs
- Locations
- Dimensions
- # of voxels
- Mass
- Volume
- Mean
- Stdev

```
[Label-num] 1
[Label-id] 1
[Slice-first] 31
[Slice-last] 243
[Row-first] 173
[Row-last] 351
[Column-first] 118
[Column-last] 387
[Dimension-x] (mm) 250.49
[Dimension-y] (mm) 166.06
[Dimension-z] (mm) 319.50
[Voxels] 769289
[Mass] (g) 1130.88
[Volume] (cc) 993.18
[Mean] (MHU) 1165.98
[Standard-deviation] (MHU) 58.88

[Label-num] 2
[Label-id] 2
[Slice-first] 44
[Slice-last] 141
[Row-first] 297
[Row-last] 341
[Column-first] 227
[Column-last] 318
[Dimension-x] (mm) 85.35
[Dimension-y] (mm) 41.75
[Dimension-z] (mm) 147.00
[Voxels] 123153
[Mass] (g) 163.72
[Volume] (cc) 159.00
[Mean] (MHU) 1054.45
[Standard-deviation] (MHU) 67.77
```

10/21/14 SW Tools & Ground Truth 32

Detection Determination Scoring (dder)

- Synopsis
  - scores an input label image produced by an ATR against the corresponding GT image based on specified values of precision and recall
  - writes results to **two log files**
- Input Files
  - ATR label image [.fits.gz]
  - GT label image [.fits.gz]
  - Object database file [.csv]
  - Packing database file [.csv]
- Output Files
  - dder summary log file [.txt] summary scoring information
  - dder false alarm log file [.xls] information for each false alarm

12/5/13 TO4 Monthly Status Call 33

Log File: dder_log_summary.txt

Summary scoring info for a single SSN
- Info for each target in GT label image
- Info for each label in ATR label image
- Scoring info
  - Detections
  - False alarms
  - Misses

NOTE: This file is read by pdpfa to compile scoring info across SSNs for calculating PD/PFA

```
Information for GT label image (targets and pseduo-targets)
-----------------------------------------------------------
[Number-of-targets] 1

[Target-num] 1
[Target-id] 6002
[Label-pixels] 197552
[Label-volume (cc)] 255.0
[Label-column-first] 126
[Label-column-last] 201
[Label-row-first] 246
[Label-row-last] 344
[Label-slice-first] 189
[Label-slice-last] 231
[Label-dimension-row (mm)] 90.9
[Label-dimension-col (mm)] 69.6
[Label-dimension-slice (mm)] 63.0
[Target-object-database-description] Breast Milk Bottle 5% Saline
[Target-material-form] bulk
[Target-material-subtype] saline
[Target-dimension-x (mm)] None Recorded
[Target-dimension-y (mm)] None Recorded
[Target-dimension-z (mm)] None Recorded
[Target-mass (g)] 253.0
[Target-volume (cc)] None Recorded
[Target-packing-database-description] Saline
[Target-xmin] 126
[Target-xmax] 200
[Target-ymin] 243
[Target-ymax] 343
[Target-zmin] 188
[Target-zmax] 231
[Target-level-of-difficulty] high
[Target-location-code] bca
[Target-orientation-code] y+

[Number-of-pseudo-targets] 0
```

2/5/2015 TO4 Monthly Status Update 34

Log File: dder_false_alarms.xls

Info about each false alarm produced by the ATR for the given SSN

- Label ID of false alarm
- Intersecting target IDs (if any)
- Form and subtype of intersecting targets
- Precision/Recall of intersecting targets

NOTE: This file is read by pdpfa to compile overall false alarm info across SSNs

	A	B	C	D	E	F	G
1	SSN	ATR_Label_ID	Intersecting_Target_ID	Intersecting_Target_Form	Intersecting_Target_Subtype	Precision	Recall
2	6	4	NA	NA	NA	NA	NA
3	6	3	7006	bulk	saline	1	0.39
4	6	2	7007	sheet	rubber	0	0
5	6	2	7006	bulk	saline	0.95	0.31

10/21/14 SW Tools & Ground Truth 35

PD/PFA Scoring (pdpfa)

- Synopsis
  - compiles the results from a specified set of images that have been scored using *dder*
  - determines PD and PFA statistics
  - writes results to **four log files**
- Input Files
  - List of log files output from dder [.txt]
  - Object database file [.csv]
  - Packing database file [.csv]
- Output Files
  - pdpfa summary log file [.txt] summary PD/PFA information
  - pdpfa detections log file [.xls] indicates whether each target was detected or missed
  - pdpfa pds log file [.xls] PD/PFA statistics
  - pdpfa false alarm log file [.xls] information for each false alarm produced

12/5/13 TO4 Monthly Status Call 36

Log File: pdpfa_log_summary.txt

Summary PD/PFA info

- PD for targets only
- PD for pseudo-targets only
- PD for targets and pseudo-targets
- PFA
- Average number false alarms

```
[Total-num-scans] 188
[Total-num-objects] 1851
[Total-num-non-targets] 1371
[Total-num-targets-and-pseudo-targets] 480
[Total-num-targets] 407
[Total-num-pseduo-targets] 73

[PFA] 0.24
[Average-num-false-alarms] 2.36

PD for targets only
-------------------
[PD-targets-overall] 0.77

[PD-targets-low-difficulty] 0.87
[PD-targets-high-difficulty] 0.73

[PD-targets-clay] 0.78
[PD-targets-rubber] 0.89
[PD-targets-saline] 0.62

[PD-targets-bulk] 0.78
[PD-targets-sheet] 0.89

PD for pseudo-targets only
--------------------------
[PD-pseudo-targets-overall] 0.34

[PD-pseudo-targets-low-difficulty] N/A
[PD-pseudo-targets-high-difficulty] 0.34

[PD-pseudo-targets-clay] 0.90
[PD-pseudo-targets-rubber] 0.30
[PD-pseudo-targets-saline] 0.63
[PD-pseudo-targets-powder] 0.83

[PD-pseudo-targets-bulk] 0.35
[PD-pseudo-targets-sheet] 0.30

PD for targets AND pseudo-targets
---------------------------------
[PD-targets-and-pseudo-targets-overall] 0.78

[PD-targets-and-pseudo-targets-low-difficulty] 0.87
[PD-targets-and-pseudo-targets-high-difficulty] 0.66

[PD-targets-and-pseudo-targets-clay] 0.79
[PD-targets-and-pseudo-targets-rubber] 0.85
[PD-targets-and-pseudo-targets-saline] 0.62
[PD-targets-and-pseudo-targets-powder] 0.83

[PD-targets-and-pseudo-targets-bulk] 0.84
[PD-targets-and-pseudo-targets-sheet] 0.85
```

2/5/2015 TO4 Monthly Status Update

Log File: pdpfa_log_detections.xls

Indicates whether each target/pseudo-target was detected

	A	B	C	D	E	F	G	H
1	SSN	Target ID	Detected	Material Type	Material Subtype	Material Form	Difficulty	Mass [g]
2	4	6002	1	t	saline	bulk	high	253
3	5	6002	0	t	saline	bulk	high	253
4	6	7006	0	t	saline	bulk	high	536
5	6	7007	1	t	rubber	sheet	high	1735
6	7	7007	1	t	rubber	sheet	high	1735
7	7	6011	0	t	saline	bulk	high	285
8	7	6012	0	t	saline	bulk	high	285
9	7	6001	1	t	saline	bulk	high	251
10	8	6002	1	t	saline	bulk	high	253
11	8	6026	0	pt	powder	bulk	high	277
12	9	6002	1	t	saline	bulk	high	253
13	9	6026	0	pt	powder	bulk	high	277
14	10	6012	1	t	saline	bulk	high	285
15	10	6001	1	t	saline	bulk	high	251

2/5/2015 TO4 Monthly Status Update 38

Log File: pdpfa_log_false_alarms.xls

Info about each false alarm produced by the ATR

- Label ID of false alarm
- Intersecting target IDs (if any)
- Form and subtype of intersecting targets
- Precision/Recall of intersecting targets

	A	B	C	D	E	F	G
1	SSN	ATR Label ID	Intersecting Target ID	Intersecting Target Form	Intersecting Target Subtype	Intersecting Target Precision	Intersecting Target Recall
2	4	1	NA	NA	NA	NA	NA
3	5	1	NA	NA	NA	NA	NA
4	6	4	NA	NA	NA	NA	NA
5	6	3	7006	bulk	saline	1	0.39
6	6	2	7007	sheet	rubber	0	0
7	6	2	7006	bulk	saline	0.95	0.31
8	7	3	NA	NA	NA	NA	NA
9	7	2	NA	NA	NA	NA	NA

10/21/14 SW Tools & Ground Truth 39

Log File: pdpfa_log_pds.xls

PD info based on

- Type
- Subtype
- Form
- LOD

PFA info

- PFA
- Average number false alarms

	A	B	C	D	E	F
1	Target Type	Target Subtype or Form	Level of Difficulty	Num Targets	Num Detected	PD [%]
2	All	All	All	480	337	70.2
3	All	Clay	All	121	96	79.3
4	All	Rubber	All	168	143	85.1
5	All	Saline	All	157	97	61.8
6	All	Powder	All	34	1	2.9
7	All	Bulk	All	333	212	63.7
8	All	Sheet	All	147	125	85
9	All	All	High	390	257	65.9
10	All	Clay	High	92	72	78.3
11	All	Rubber	High	135	110	81.5
12	All	Saline	High	129	74	57.4
13	All	Powder	High	34	1	2.9
14	All	Bulk	High	264	153	58
15	All	Sheet	High	126	104	82.5
16	Target	All	All	407	312	76.7
17	Target	Clay	All	111	87	78.4
18	Target	Rubber	All	158	140	88.6
19	Target	Saline	All	138	85	61.6
20	Target	Bulk	All	270	190	70.4
21	Target	Sheet	All	137	122	89.1
22	Target	All	Low	77	67	87
23	Target	Clay	Low	29	24	82.8
24	Target	Rubber	Low	22	22	100
25	Target	Saline	Low	26	21	80.8
26	Target	Bulk	Low	56	46	82.1
27	Target	Sheet	Low	21	21	100
28	Target	All	High	317	232	73.2
29	Target	Clay	High	82	63	76.8
30	Target	Rubber	High	125	107	85.6
31	Target	Saline	High	110	62	56.4
32	Target	Bulk	High	201	131	65.2
33	Target	Sheet	High	116	101	87.1
34	Pseudo-target	All	High	73	25	34.2
35	Pseudo-target	Clay	High	10	9	90
36	Pseudo-target	Rubber	High	10	3	30
37	Pseudo-target	Saline	High	19	12	63.2
38	Pseudo-target	Powder	High	34	1	2.9
39	Pseudo-target	Bulk	High	63	22	34.9
40	Pseudo-target	Sheet	High	10	3	30
41						
42	Num Non-targets	Num FAs	PFA [%]			
43	1371	330	24.1			
44						
45	Num FAs	Num Scans with FAs	Avg Num FAs			
46	330	140	2.36			

2/5/2015 TO4 Monthly Statu

Gen_pdpfa

- Synopsis
  - Runs a specified ATR on a specified set of CT images to produce a set of corresponding label images
  - Scores the label images against the ground truth (GT) label images using *dder*
  - Generates PD/PFA statistics using *pdpfa*

- Input Files
  - CT image(s) [.fits.gz]
  - GT image(s) [.fits.gz]
  - Label image(s) [.fits.gz]
  - Object database file [.csv]
  - Packing database file [.csv]
  - Height Database file [.csv]
  - List of SSNs for which to calculate PD and PFA statistics

- Output Files
  - All log files generated by *satr, dder,* and pdpfa
  - dder log list file [.txt]   list of *dder* log files used by *pdpfa* to compile results
  - gen_pdpfa.sh log file [.txt]   standard output from *satr*, *dder*, and *pdpfa*

10/21/14   SW Tools & Ground Truth   41

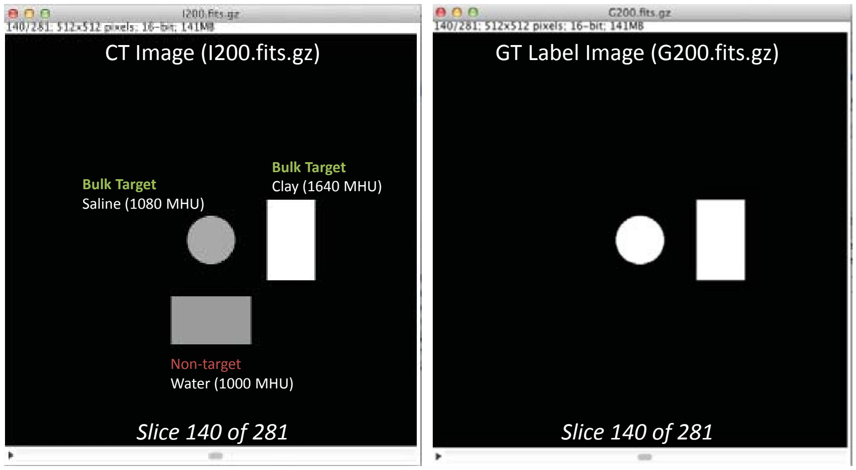
g3d Test Images

- Two test images and corresponding GT label images are available
  - One w/ bulk targets, one w/ sheet targets
  - All objects have realistic MHU values
  - Test images & documentation located on FTP site:
    - eng_research_TO4/test_images

CT Image Name	GT Label Image Name	Image Dimensions [x y z]	# Objects	# Targets	Contents
I200.fits.gz	G200.fits.gz	[512 512 281]	3	2	1) Saline bulk target 2) Clay bulk target 3) Water non-target
I201.fits.gz	G201.fits.gz	[512 512 281]	3	2	1) Rubber sheet target 2) Rubber sheet target 3) Water non-target

04/03/14   TO4 Monthly Status Update   42

g3d Test Image 1



04/03/14 TO4 Monthly Status Update 43

g3d Test Image 2

04/03/14 TO4 Monthly Status Update 44

Conclusions

TO4 Database

- Excel spreadsheet created during scanning
- Detailed information for all scanned objects and SSNs

Ground Truth Development and Label Images

- Semi-automated segmentation program for creating GT label images from CT images
- GT label images available for all TO4 CT images (188 total)

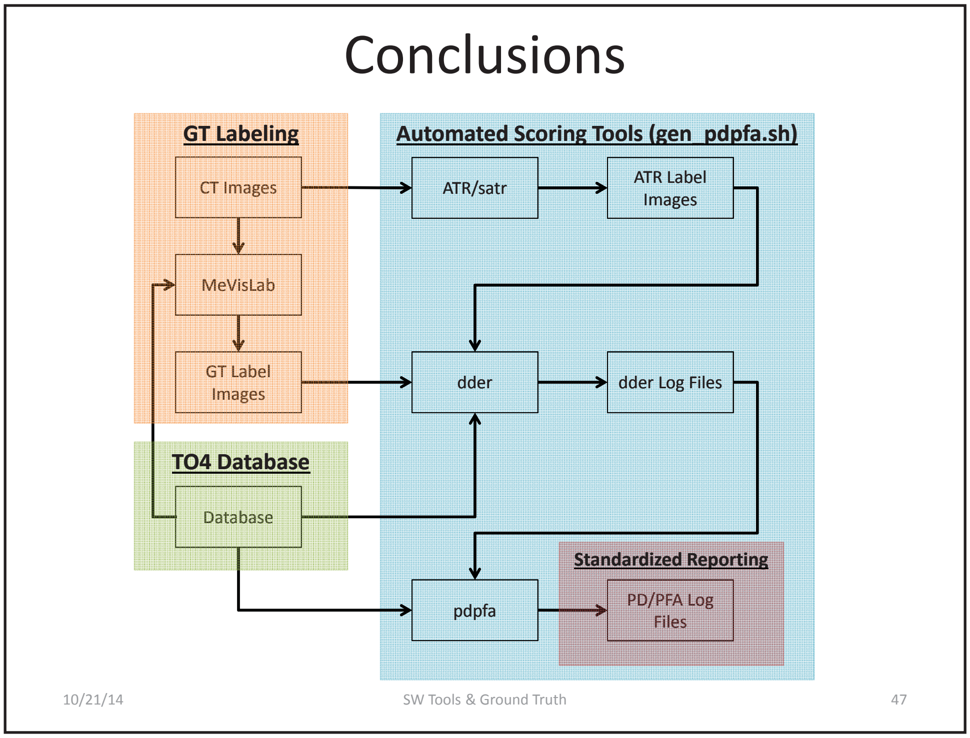
10/21/14 SW Tools & Ground Truth 45

Conclusions

Automated Scoring Tools

- Offload development of scoring tools from researchers
- Provide consistent method of scoring across ATRs
  - PD/PFA
- Provide consistent reporting via generated log files
  - Overall PD/PFA
  - PD per type, sub-type, form, and LOD
  - Detection status for each target/pseudo-target
- Facilitate improvement of ATR performance
  - Tweaking of scoring parameters (precision, recall)
  - Useful information in log files
  - Reporting of incomplete detections

10/21/14 SW Tools & Ground Truth 46


Conclusions
GT Labeling
CT Images
MeVisLab
GT Label Images
TO4 Database
Database
Automated Scoring Tools (gen_pdpfa.sh)
ATR/satr
ATR Label Images
dder
dder Log Files
pdpfa
Standardized Reporting
PD/PFA Log Files
10/21/14
SW Tools & Ground Truth
47

11.5.2.3 ATR Development – Ye

"Automatic Target Recognition (ATR) for CT Luggage Screening"

Automatic Target Recognition (ATR) for CT Luggage Screening

Dong Hye Ye, Pengchong Jin, Charles Bouman
School of Electrical and Computer Engineering, Purdue University
12/15/2014

Institution and Researchers

- Dong Hye Ye – Post-doctoral Researcher of ECE, Purdue University. BSEE SNU, Ph.D. U. Penn (2013). Machine learning for image segmentation, registration and classification
- Pengchong Jin – 6th year Ph.D. student, Purdue University. B.Eng, ECE from HKUST (2009). Statistical signal processing, inverse problems
- Charles Bouman – Showalter Professor of ECE, Purdue University. BSEE U. Penn, Ph.D. Princeton (1989). Stochastic image modeling, image rendering, tomography

2

PD/PFA results

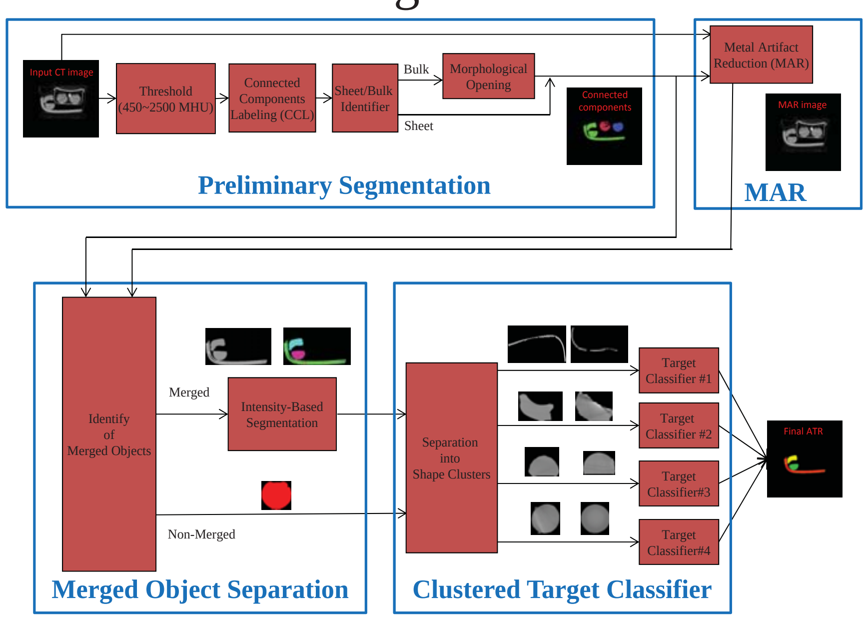
Target Type	Target Subtype	Level of Difficulty	Num Targets	Num Detected	PD [%]
Target	**All**	**All**	**407**	**387**	**95.1**
Target	Clay	All	111	106	95.5
Target	Rubber	All	158	151	95.6
Target	Saline	All	138	130	94.2
Target	Bulk	All	270	260	96.3
Target	Sheet	All	137	127	92.7
Target	All	Low	77	75	97.4
Target	Clay	Low	29	28	96.6
Target	Rubber	Low	22	21	95.5
Target	Saline	Low	26	26	100
Target	Bulk	Low	56	54	96.4
Target	Sheet	Low	21	21	100
Target	All	High	317	299	94.3
Target	Clay	High	82	78	95.1
Target	Rubber	High	125	119	95.2
Target	Saline	High	110	102	92.7
Target	Bulk	High	201	193	96
Target	Sheet	High	116	106	91.4
Pseudo-target	Sheet	High	10	9	90

Num Non-targets	Num FAs	PFA [%]
1371	**110**	**8**

Clay, Rubber, Saline
Bulk > Sheet

3

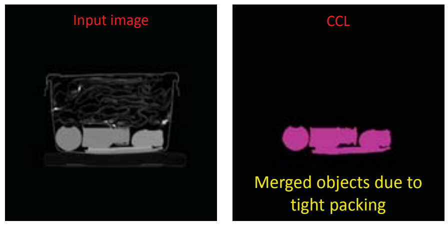
ATR High Level



4

Overview: Preliminary Segmentation

- Threshold out objects 450-2500 MHU
- CCL segmentation
  - Create merged objects

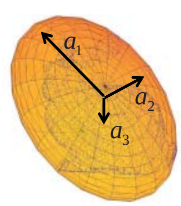


- Morphological processing to separate components
  1. Apply sheet/bulk identifiers
  2. Perform morphological opening on bulk structures

5

Sheet/Bulk Identifier: Shape Features

1. Intensity: Min, Max, Mean
2. Physical: Mass, Volume, Surface area
3. Histogram: Number of peaks in histogram
4. Shape: Minimum volume enclosing ellipsoid [P. Kumar, 2003]
   - Ellipsoid Axes : a_1, a_2, a_3
   - Axis Ratio: $\min\{a_1, a_2, a_3\} / \max\{a_1, a_2, a_3\}$
   - Volume Ratio: object volume / ellipsoid volume



- Example

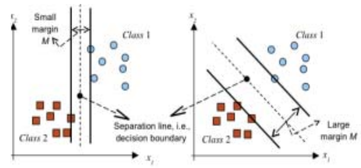
Small Axis Ratio
0.1425

Large Axis Ratio
0.5879

6

Sheet/Bulk Identifier: Supervised Learning

- Support Vector Machine (SVM) [C. Cortes, 1995]
  - Find decision boundary that separates two groups while maximizing margins of support vectors

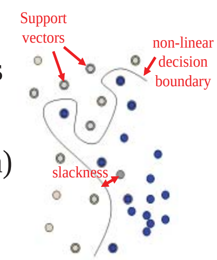


- Training on 485 Ground-Truth objects
  - 5-Fold Cross Validation (CV) for parameters
    1. Slack variable
    2. Gaussian kernel (Radial Basis Function)

Gaussian kernel   Input features

$$\phi(f, f') = \exp\left(-\frac{\| f - f' \|^2}{2\sigma^2}\right)$$

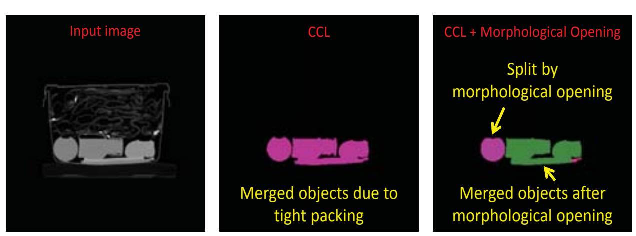
Kernel size



7

Morphological Opening on Bulk Structures

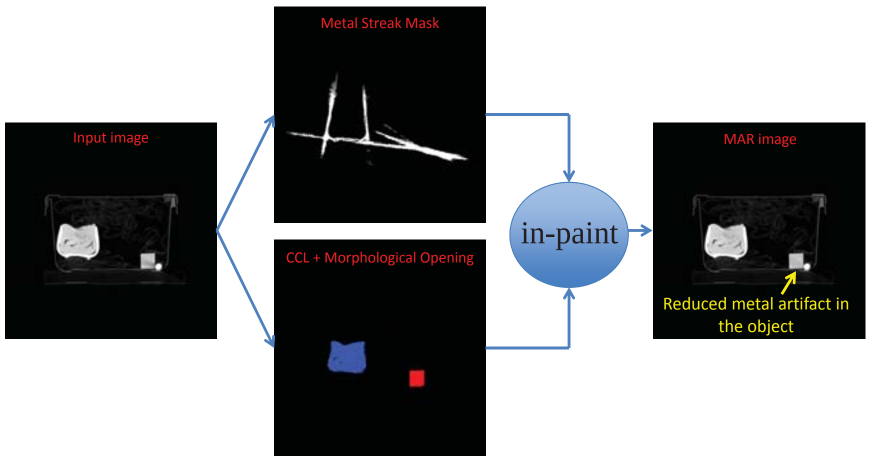
- Morphological opening separates bulk structures
- Morphological opening
  - Erosion followed by dilation



8

Overview: Metal Artifact Reduction

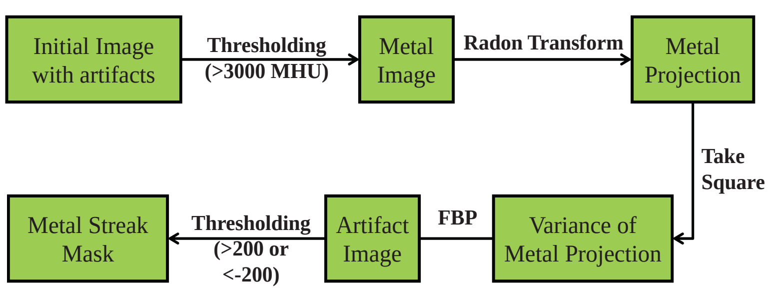
- Detect metal streak mask
- In-paint streak mask region



9

Metal Streak Mask Detection

- Algorithm for mask detection [S. Karimi, 2014]
  - Estimate variance due to metal in sinogram
  - Reconstruct steak mask

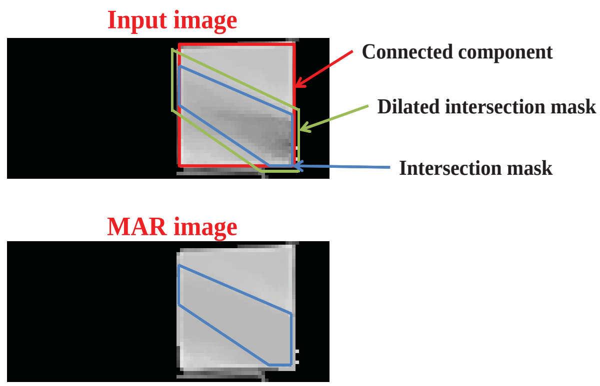


10

In-painting in Metal Streak Mask Region

For each connected component

1. Perform intersection with the metal streak mask
2. Morphologically dilate the intersection mask
3. Replace CT values in intersection with intensity of nearby voxels



11

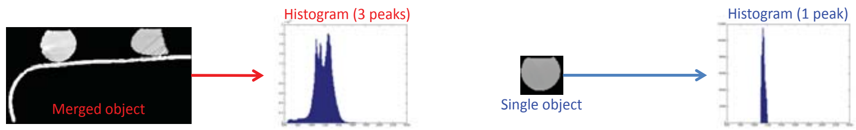
Overview: Merged Object Separation

- Split processing to separate merged objects
  - Identify merged objects
  - Apply intensity-based segmentation to merged objects

12

Identify Merged Objects

- Same shape features as used for sheet/bulk identifier



- Same supervised learning for sheet/bulk identifier
  - SVM with Gaussian kernel (5-fold CV for parameters)
- Leave-one-bag out testing

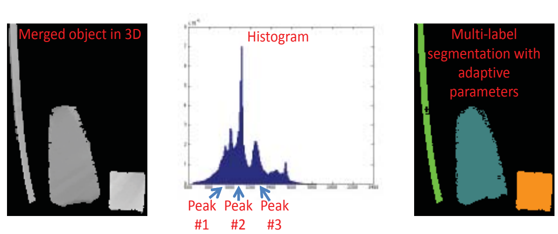
Training objects in remaining 187 bags (Around 1087 objects) → Shape Feature Extraction → Train SVM on Merged/Non-Merged

Testing objects in one bag → Shape Feature Extraction → Apply trained Merged/Non-Merged SVM → Merged/ Non-Merged

13

Intensity-Based Segmentation: Parameters

- Two sensitive parameters
  - Number of materials: n
  - Mean intensity for material k : I^k
- Adaptive parameter setting
  1. Find major peaks in normalized histogram ($n \le 3$)
  2. Assign histogram bins at major peaks to I^k



14

Intensity-Based Segmentation: Formulation

- Multi-label segmentation [J. Yuan, 2010]

$$\min_{0 \le u_j^k \le 1} \sum_{k=1}^{n} \sum_{j} u_j^k \left| x_j - l^k \right|^2 + \alpha \sum_{k=1}^{n} \sum_{j} \left| \nabla u_j^k \right|$$

Probability of voxel *j* being assigned to material *k*

Image intensity at voxel *j*

Total-variation (TV) regularization

- Cost function
  - First term: Intensity-Based fidelity
  - Second term: Label smoothness

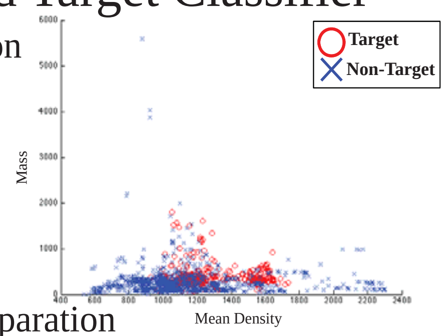
- Assign the voxel j to k^* where u_j^k is the maximum

$$k^* = \arg\max_k u_j^k$$

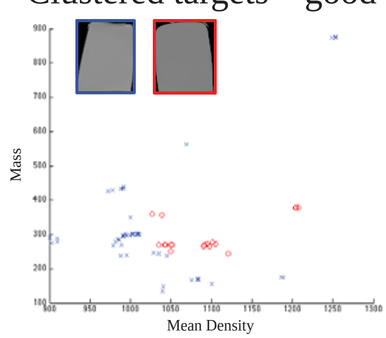
15

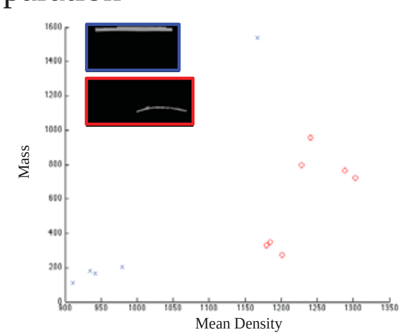
Overview: Clustered Target Classifier

- All targets – poor separation



- Clustered targets – good separation

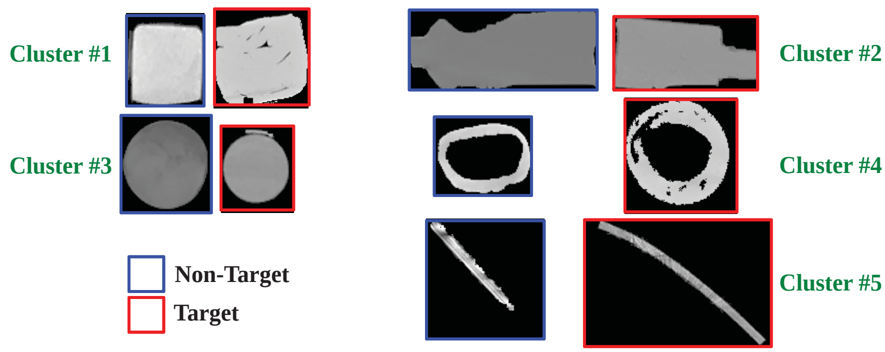




16

Shape Clustering of Training Set

- K-means clustering
  - Same shape features as used for sheet/bulk identifier
  - 15 shape clusters (empirically chosen based on dataset)
- Examples of shape clusters

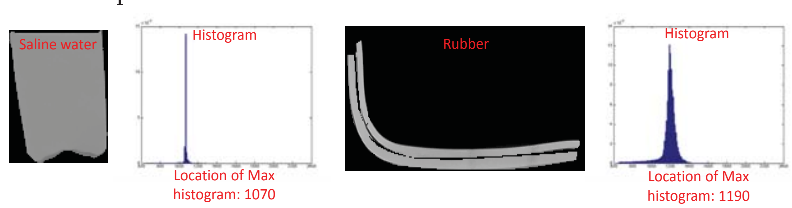


17

Target Classifier: Histogram Features

1. Intensity: Min, Max, Mean
2. Physical: Mass
3. Histogram (bin size: [450:10:2500])
   - Location of max histogram
   - Normalized histogram: 206-dimensional vector

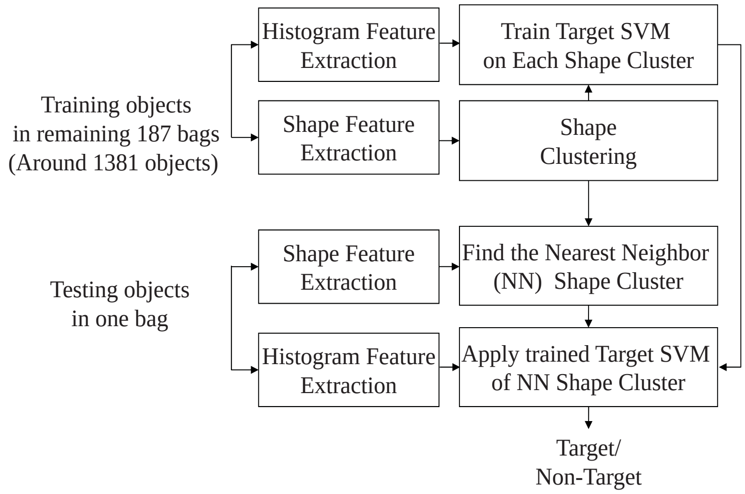
- Feature Selection: min-Redundancy Max-Relevance [H. Peng, 2005]
  - Avoid curse of dimensionality issue
  - Example of selected feature



18

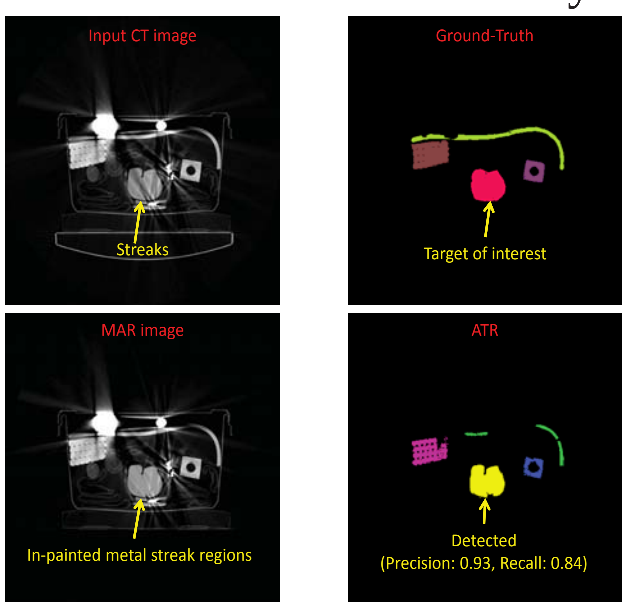
Target Classifier: Supervised Learning

- SVM with Gaussian kernel (5-fold CV for parameters)
- Leave-one-bag out testing



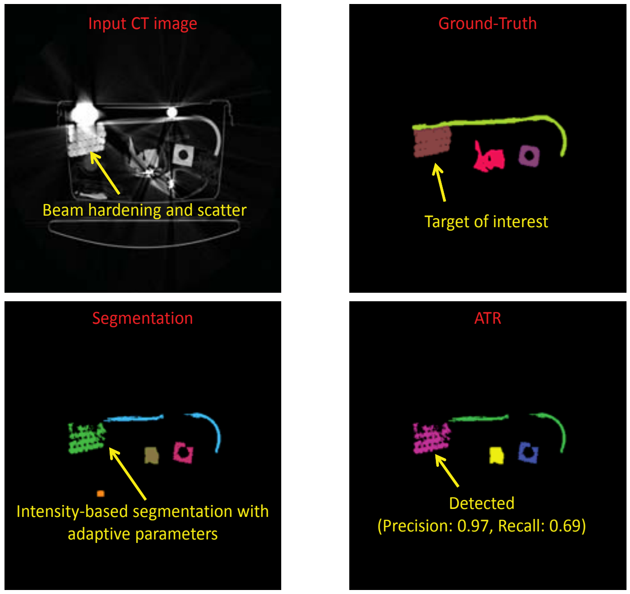
19

Example (1) - SSN 13 (Slice 105)
Bulk with bad streaks caused by metal



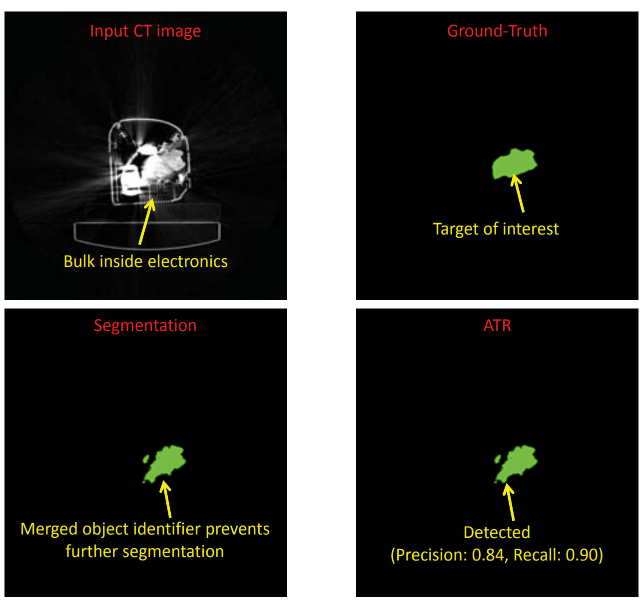
20

Example (2) -SSN 13 (Slice 128)
Bulk with beam hardening and scatter



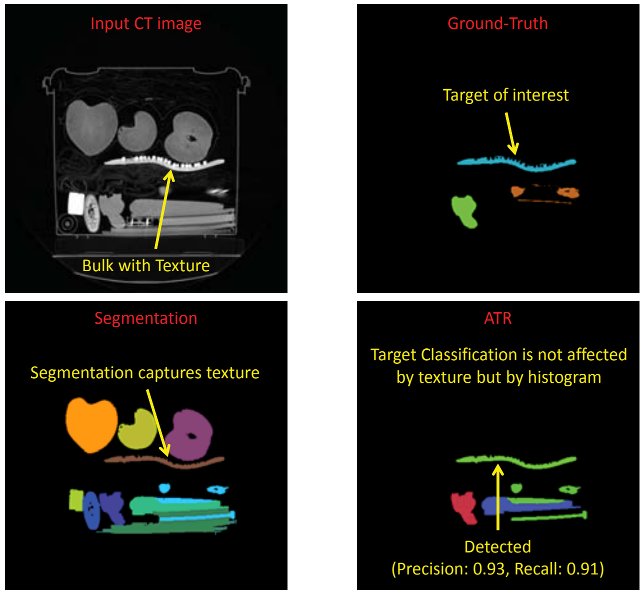
21

Example (3) - SSN 35 (Slice 49)
Bulk inside electronics



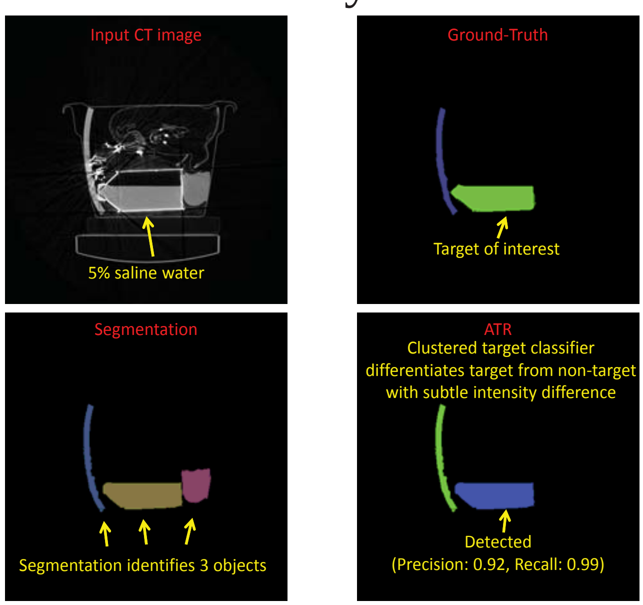
22

Example (4) - SSN 193 (Slice 198)
Bulk with texture



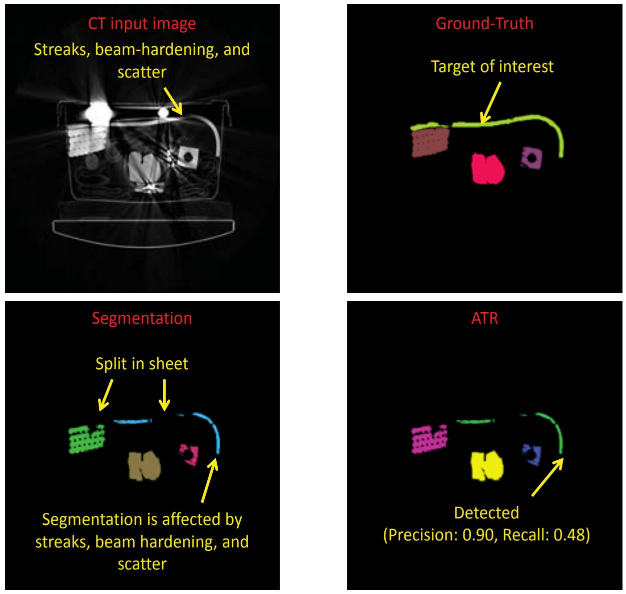
23

Example (5) - SSN 63 (Slice 45)
Bulk with density close to water



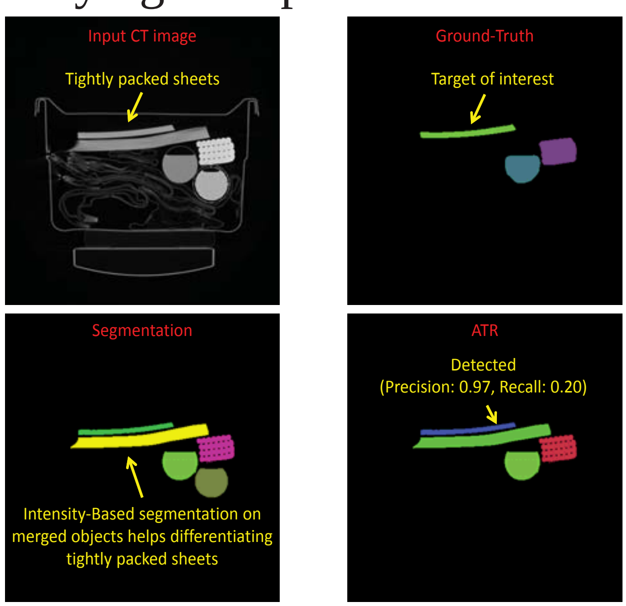
24

Example (6) - SSN 13 (Slice 111)
Sheet with beam-hardening and scatter



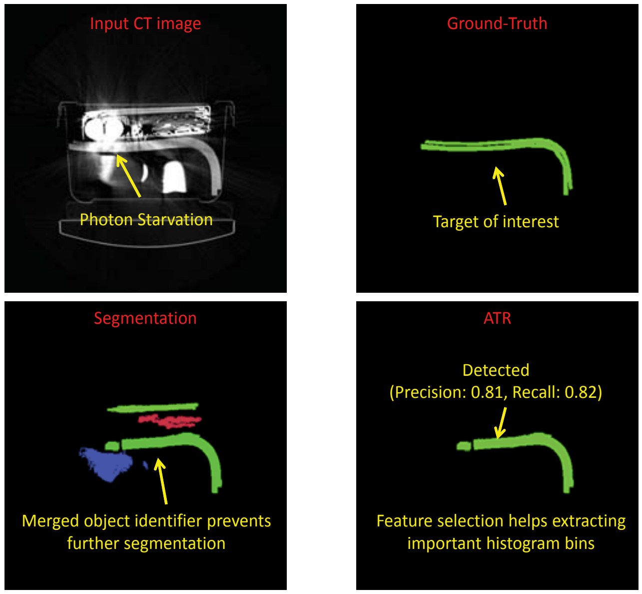
25

Example (7) - SSN 33 (Slice 46)
Sheet laying on top of another flat object



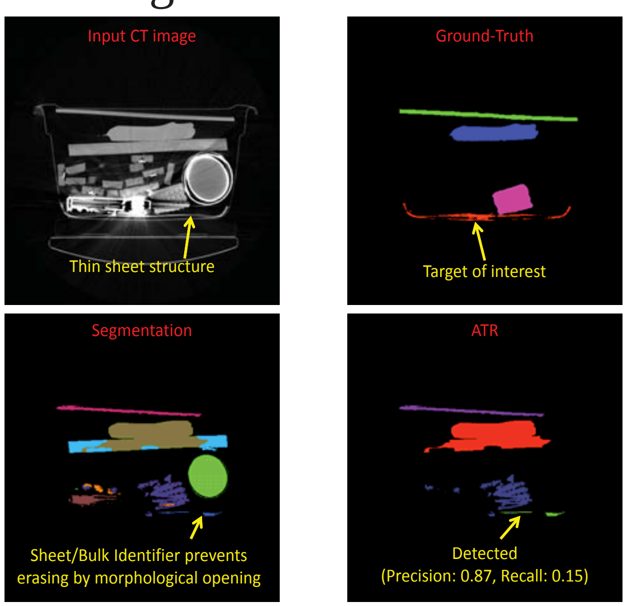
26

Example (8) - SSN 11 (Slice 94)
Object with lots of photon starvation



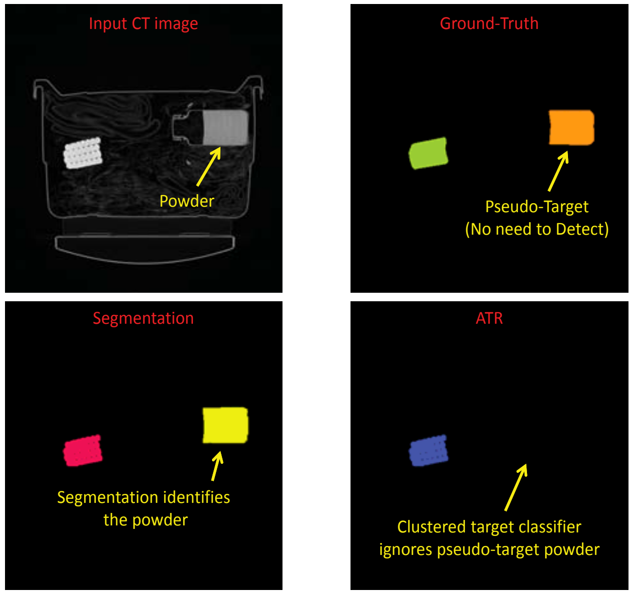
27

Example (9) - SSN 18 (Slice 125)
Pseudo-target sheet based on thickness



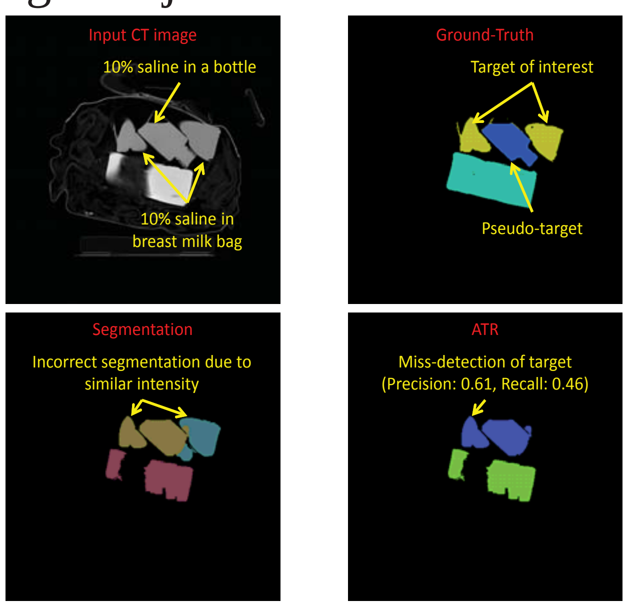
28

Example (10) - SSN 12 (Slice 105): Pseudo-target powder based on density



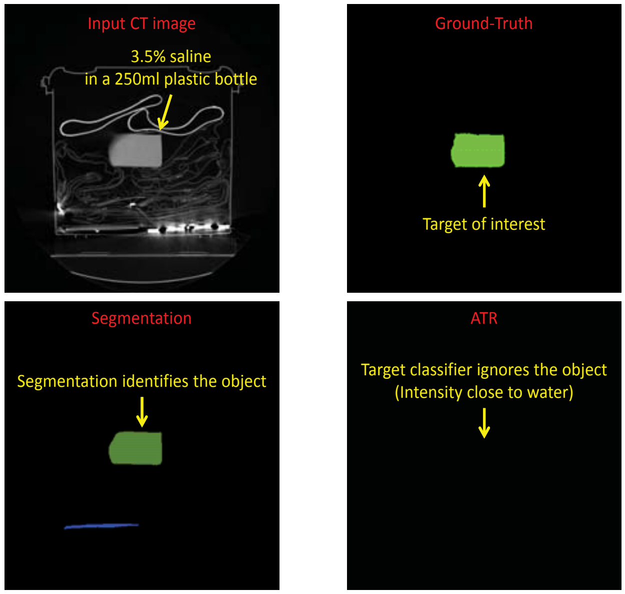
29

Missed Target (1) - SSN 17 (Slice 50) Merged objects with similar intensity



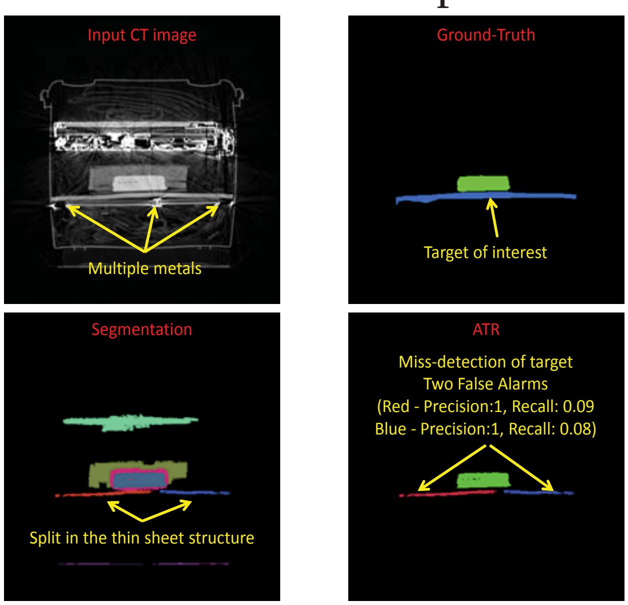
30

Missed Target (2) - SSN 91 (Slice 125)
Low-density saline



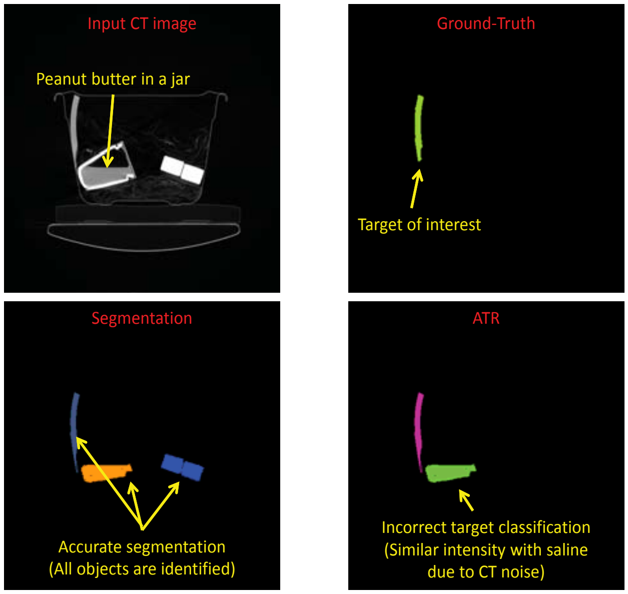
31

False Alarm (1) - SSN 73 (Slice 100)
Thin sheet with multiple metals



32

False Alarm (2) - SSN 52 (Slice 140)
Object in a metal container



33

Strengths

- Segmentation
  - Preserves thin structures
  - Separates merged objects
- Metal Artifact Reduction
  - Corrects CT values inside streak mask
- Target Classification
  - Training with clustered targets reduces overlap in features
  - Higher classification accuracy

34

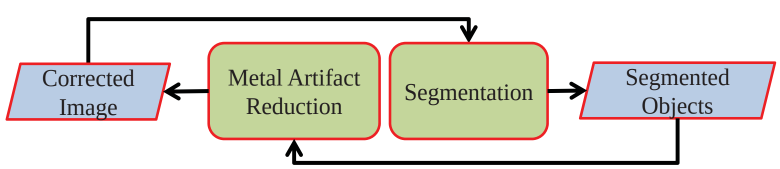
Weaknesses

- Segmentation
  - Metal artifacts still can split sheet structures
  - Intensity-Based segmentation can not separate merged objects with the same material
- Metal Artifact Reduction
  - In-painting can produce artificial CT values when the streak mask is not accurate
- Target Classification
  - Missing very low-density saline targets (close to water)
  - CT noise can produce false alarms (peanut butter in a metal container)

35

Proposed Improvement

- Joint Metal Artifact Reduction and Segmentation
  - Less splitting by metal artifact reduction
  - Accurate in-painting with the help of segmentation



Original Image | Corrected Image | Segmentation

36

Comments on the data and the process

- CT images
  - Need scans with no target in the dataset
- Ground-Truth (GT) Labels
  - Need GT labels of non-target objects for evaluation of segmentation and training of target classifier
- Scoring Tool
  - Very helpful for automatic evaluation
- Evaluation
  - Require a single combined metric for PD and PFA due to trade-off (i.e. weighted sum)

37

Summary

- Main challenges in ATR
  - Metal artifacts and tight packing
- Segmentation
  - Morphological opening for bulk structures
  - Metal artifact detection and in-painting in image domain
  - Intensity-based segmentation for merged objects
- Target Classification
  - Shape clustering of training dataset
- Evaluation
  - PD: 0.95, PFA: 0.08

38

Back-Up

39

Shape Features

- Intensity: Min ($\min_{j \in c^k} x_j$), Max ($\max_{j \in c^k} x_j$), Mean ($\sum_{j \in c^k} x_j / \sum_{j \in c^k} 1$)
- Physical: Mass ($\sum_{j \in c^k} x_j$), Volume ($\sum_{j \in c^k} 1$), Surface Area ($\sum_{j \in \partial c^k} 1$)
- Minimum volume enclosing ellipsoid

$$\begin{aligned} &\min \ \log(\det(D)) \\ &\text{s.t.} \quad (P_i - d)^T D (P_i - d) \le 1 \end{aligned}$$

$P \in \mathbf{R}^{3 \times N}$: coordinates in the object, D $\in \mathbf{R}^{3\times3}$: ellipse equation, d $\in \mathbf{R}^3$: center of ellipse

$$\{\lambda_1, \lambda_2, \lambda_3\} = \mathrm{eig}(D)$$

  - Ellipsoid Axes : $\{2/\sqrt{\lambda_1}, 2/\sqrt{\lambda_2}, 2/\sqrt{\lambda_3}\}$
  - Axis Ratio: $\min\{2/\sqrt{\lambda_1}, 2/\sqrt{\lambda_2}, 2/\sqrt{\lambda_3}\} / \max\{2/\sqrt{\lambda_1}, 2/\sqrt{\lambda_2}, 2/\sqrt{\lambda_3}\}$
  - Volume Ratio: $8 \cdot \sum_{j \in c^k} 1 / \sqrt{\lambda_1 \lambda_2 \lambda_3}$

- Histogram: Number of peaks in histogram (n)

40

Histogram Features

- Intensity: Min ($\min_{j \in c^k} x_j$), Max ($\max_{j \in c^k} x_j$), Mean ($\sum_{j \in c^k} x_j / \sum_{j \in c^k} 1$)
- Physical: Mass ($\sum_{j \in c^k} x_j$)
- Histogram (bin size: [450:10:2500])
  - Normalized histogram: 206-dimensional vector
  - Location of Max Histogram
- Feature Selection
  - Avoid curse of dimensionality issue
  - Cross-validate the number of selected features ([10:5:30])
  - min-Redundancy Max-Relevance (mRMR) [H. Peng, 2005]

$$\max_F \underbrace{\frac{1}{m} \sum_{f_i \in F} M(f_i, y)}_{\text{Maximum Relevance}} - \underbrace{\frac{1}{m^2} \sum_{f_i, f_j \in F} M(f_i, f_j)}_{\text{Minimum Redundancy}}$$

f_i : i^{th}-dimensional vector in feature set F, y : ground-truth label vector of training set

$M(\cdot,\cdot)$: Mutual information between two vectors, m : cardinality of F

41

Intensity-based Segmentation: Formulation

- Multi-label segmentation on discrete setting

$$\min_{u_j^k \in \{0,1\}} \sum_{k=1}^{n} \sum_{j} u_j^k \underbrace{\left| x_j - l^k \right|^2}_{\text{Intensity-based Fidelity}} + \alpha \sum_{k=1}^{n} \sum_{j} \underbrace{| \nabla u_j^k |}_{\text{Label Smoothness}}$$

u_j^k : binary indicator for assigning voxel j to material k (label intensity: l^k)

x_j : image itensity at j, α: regularization parameter

- Convex Relaxation [J. Yuan, 2010]

$$\min_{u_j^k} \sum_{k=1}^{n} \sum_{j} u_j^k \left| x_j - l^k \right|^2 + \alpha \sum_{k=1}^{n} \sum_{j} | \nabla u_j^k |$$

$$\text{s.t.} \sum_{k=1}^{n} u_j^k = 1;\ u_j^k \in [0,1]$$

Relax the binary configuration to the probability

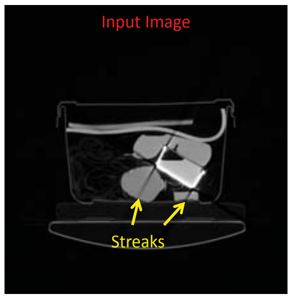
- Assign the voxel to k^* where u_j^k is the maximum

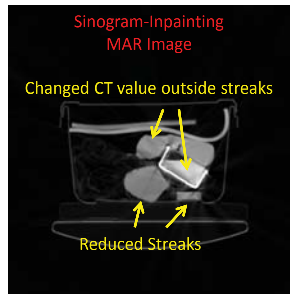
$$\mathrm{k}^* = \arg\max_k u_j^k$$

42

MAR: Existing Method

- Sinogram-Inpainting [H. Tuy, 1993]
  - Identify metal traces in original sinogram
  - Interpolate sinogram in metal traces
  - Reconstruct the interpolated sinogram
- Limitation
  - Secondary artifacts outside streaks





43

In-painting in Metal Artifacts Mask

- Strategy
  - Correct CT values only inside the objects
- Algorithm

For each CCL results $C = \{c^1, \ldots, c^r, \ldots, c^R\}$

1. Perform intersection with the metal artifact mask B

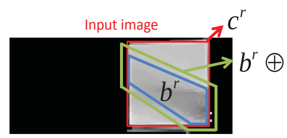
$$b^r = B \cap c^r$$

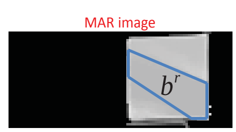
2. Replace the value in b^r with mean intensity of nearby voxels in c^r

Morphological Dilation

$$N(b^r) = (b^r \oplus) - b^r$$

$$x_{j \in b^r} = \frac{\sum_{i \in c^r \cap N(b^r)} x_i}{\sum_{i \in c^r \cap N(b^r)} 1}$$





44

Metal Artifact Detection

- Assumption
  - Variances in metal projections cause artifacts
- Algorithm
  1. Calculate metal projection $p_{metal} \in \mathbf{R}^M$

$$p_{metal,\,i} = \sum_{j=1}^{N} A_{i,j} x_j \delta(x_j \geq T_{metal}),\ T_{metal} = 3000$$

$$x \in \mathbf{R}^N : \text{Input Image},\ A \in \mathbf{R}^{M \times N} : \text{ Forward Matrix}$$

  2. Model variance of metal projection $h : \mathbf{R}^M \rightarrow \mathbf{R}^M$

$$h(p_{metal,\,i}) = p^2_{\ metal,\,i}$$

  3. Relate metal artifact $z \in \mathbf{R}^N$ with the modeled variance

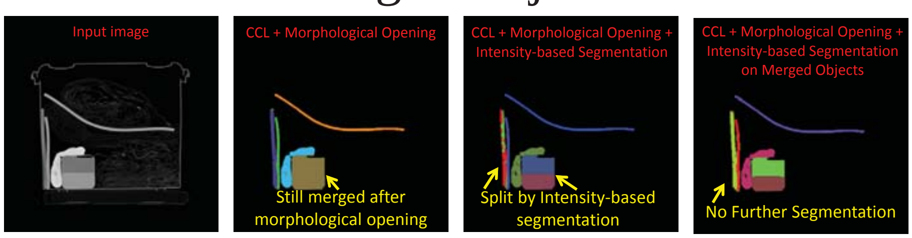
$$z_j = \sum_{i=1}^{M} A^T_{j,i} h(p_{metal,\,i}) = \sum_{i=1}^{M} A^T_{j,i} p^2_{metal,\,i}$$

  4. Create the metal artifact mask $b \in \{0,1\}^N$

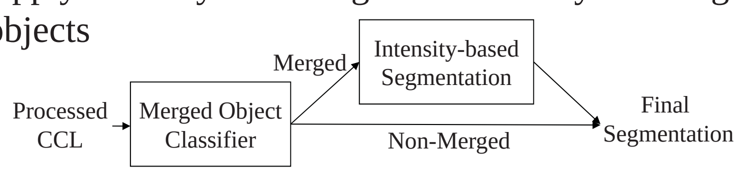
$$b_j = \begin{cases} 1, & |z_j| > T_{artifact} \text{ and } x_j < T_{metal} \\ 0, & |z_j| \leq T_{artifact} \text{ or } x_j \geq T_{metal} \end{cases},\ T_{artifact} = 200$$

45

Intensity-based Segmentation on Merged Objects



- Need to further segment after morphological opening
- Use intensity information for segmentation
  - Issue: Over-segmentation
- Apply intensity-based segmentation only for merged objects



PD/PFA results (revisited)

Target Type	Target Subtype	Level of Difficulty	Num Targets	Num Detected	PD [%]
Target	**All**	**All**	**407**	**387**	**95.1**
Target	Clay	All	111	106	95.5
Target	Rubber	All	158	151	95.6
Target	Saline	All	138	130	94.2
Target	Bulk	All	270	260	96.3
Target	Sheet	All	137	127	92.7
Target	All	Low	77	75	97.4
Target	Clay	Low	29	28	96.6
Target	Rubber	Low	22	21	95.5
Target	Saline	Low	26	26	100
Target	Bulk	Low	56	54	96.4
Target	Sheet	Low	21	21	100
Target	All	High	317	299	94.3
Target	Clay	High	82	78	95.1
Target	Rubber	High	125	119	95.2
Target	Saline	High	110	102	92.7
Target	Bulk	High	201	193	96
Target	Sheet	High	116	106	91.4
Pseudo-target	Sheet	High	10	9	90

Num Non-targets	Num FAs	PFA [%]
1371	**110**	**8**

47

Evaluation of Segmentation

	CCL	CCL + Morphological Opening	CCL + Morphological Opening + Intensity-based Segmentation
PD-overall	0.63	0.89	**0.97**
PD-overall low-difficulty	0.63	0.88	**0.99**
PD-overall high-difficulty	0.63	0.89	**0.96**
PD-overall-bulk	0.58	0.87	**0.97**
PD-overall-sheet	0.75	0.94	**0.96**
PD-pseudo target-sheet	0.70	1.00	**1.00**

- Sheet/Bulk identifier helps increasing PD-pseudo target-sheet by not applying morphological opening for thin sheets.
- Morphological opening improves PD over CCL by splitting merged objects due to tight packing.
- Intensity-based segmentation further increases PD by dividing merged objects with different materials.

48

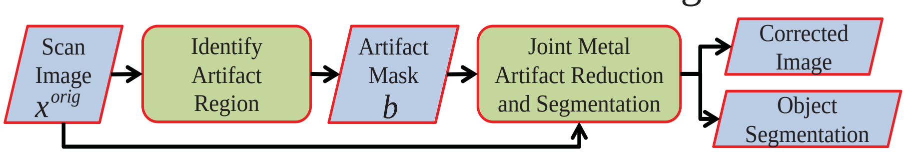
Evaluation of Target Classification

	Segmentation	Segmentation + Target Classifier One for All Training	Segmentation + Target Classifier Shape Clustering
PD-overall	**0.97**	0.89	0.95
PD-overall low-difficulty	**0.99**	0.89	0.97
PD-overall high-difficulty	**0.96**	0.89	0.94
PD-pseudo target-sheet	**1.00**	0.70	0.90
PFA-overall	0.73	0.16	**0.08**
Average num. false alarms	5.44	1.78	**1.29**

- Segmentation has high PD and high PFA.
- Target classifier significantly decreases PFA but lose the target detection due to mis-classification.
- Shape clustering improves classification performance (Higher PD and Lower PFA).

49

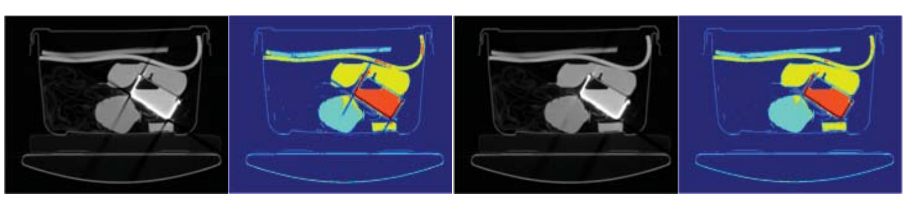
Joint Metal Artifact Reduction and Segmentation



minimize over $\{x,\nu\}$, u and μ, given a global trained dictionary Φ

$$\frac{1}{2}\sum_{i=1}^{N}(1-b_i)(x_i-x_i^{\mathrm{orig}})^2+\frac{\alpha}{2}\sum_{i=1}^{N}\|R_i x-\Phi\nu_i\|_2^2+\sum_{i=1}^{N}\gamma_i\|\nu_i\|_1+\frac{\beta}{2}\sum_{i=1}^{N}\sum_{k=1}^{K}u_{i,k}(x_i-\mu_k)^2+\lambda\sum_{i=1}^{N}\sum_{k=1}^{K}|\nabla u_{i,k}|$$

subject to $0 \le u_{i,k} \le 1, \quad \sum_{k=1}^{K} u_{i,k} = 1$



Original Image | Segmentation on Original Image | Corrected Image | Segmentation on Corrected Image

50

11.5.2.4 ATR Development – Gregor

"ALERT Task Order 4: ATR Development using Graph Algorithms"

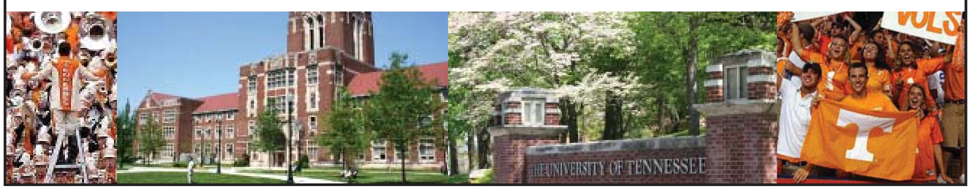
THE UNIVERSITY of TENNESSEE UT KNOXVILLE

ALERT Task Order 4
ATR Development using Graph Algorithms

Jens Gregor

Electrical Engineering and Computer Science
University of Tennessee, Knoxville, TN
jgregor@eecs.utk.edu

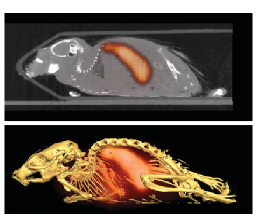
This material is based upon work supported by the U.S. Department of Homeland Security, Science and Technology Directorate, under Task Order Number HSHQDC-12-J-00056. The views and conclusions contained in this document are those of the authors and should not be interpreted as necessarily representing the official policies, either expressed or implied, of the U.S. Department of Homeland Security.



Dr. Jens Gregor

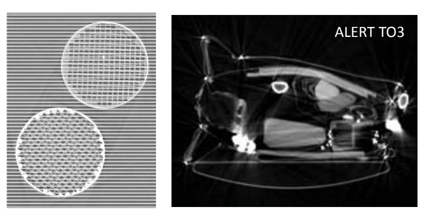
Medical imaging

PET, uSPECT, uCT



NDT imaging (incl. luggage)

Neutron, X-ray CT



- Image reconstruction, image/data analysis, parallel computing
- Academic proof-of-principle and commercial/production code

- 20+ years of experience teaching computer science at UTK
- US citizen

11/6/2014 Jens Gregor 2

PD/PFA results

Num Non-targets	Num FAs	PFA [%]
1371	274	20

Target Type	Target Subtype	Level of Difficulty	Num Targets	Num Detected	PD [%]
● **Target**	**All**	**All**	**407**	**359**	**88.2**
Target	Clay	All	111	95	85.6
Target	Rubber	All	158	148	93.7
Target	Saline	All	138	116	84.1
Target	Bulk	All	270	229	84.8
● **Target**	**Sheet**	**All**	**137**	**130**	**94.9**
Target	All	Low	77	68	88.3
Target	Clay	Low	29	26	89.7
Target	Rubber	Low	22	21	95.5
Target	Saline	Low	26	21	80.8
Target	Bulk	Low	56	47	83.9
Target	Sheet	Low	21	21	100
Target	All	High	317	279	88
Target	Clay	High	82	69	84.1
Target	Rubber	High	125	117	93.6
Target	Saline	High	110	93	84.5
Target	Bulk	High	201	170	84.6
Target	Sheet	High	116	109	94
● **Pseudo Target**	**Sheet**	**High**	**10**	**7**	**70**

11/6/2014 Jens Gregor 3

ATR Overview

Connected Component Labeling
Thresholding: 900-2200. Labeling. Merging of fragments.

Shape-based Object Splitting
Separation of sheets, bulk. Label images viewed as 3D graphs.

Density-based Object Splitting
Separation of materials using Gaussian mixture models.

Feature Extraction/Classification
Simple features. Support vector machines: sheets, bulk.

Basic approach similar to prior art. Difference is wrt underlying methods.

11/6/2014 Jens Gregor 4

ATR Design Philosophy

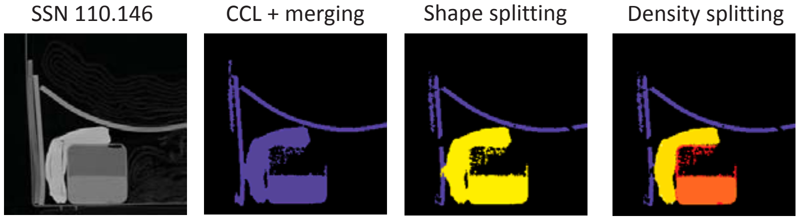
- High PD implies high recall and precision which necessitates accurate segmentation of objects
  - Graphs model spatial relations including how pixels form objects which may be used to indicate merging of sheets and bulk objects and thus facilitate separation thereof
  - Gaussian mixture models of density histograms may allow separation of objects made from different materials
- Low PFA done by discarding non-target objects
  - Use of simple, physical features ensures generality
  - Support vector machines are linear classifiers that map data to/operate in non-linear kernel space (eg pattern similarity)

11/6/2014 Jens Gregor 5

Object Splitting: Motivation

- Connected component labeling creates objects
  - Neighboring pixels grouped without object context
  - Disjoint set built for $900 \leq f \leq 2200$ and $|f_i - f_j| < 75$
- Adjacent objects merged to limit fragmentation
  - Based on proximity and simple statistics, not shape
- Some objects become fused during the process

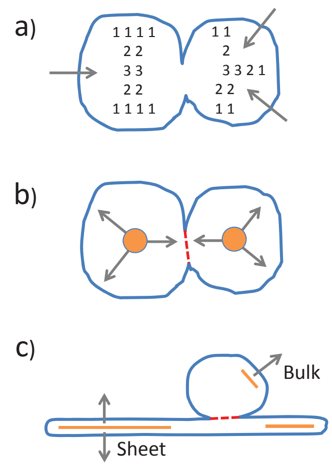
SSN 110.146 | CCL + merging | Shape splitting | Density splitting



11/6/2014 Jens Gregor 6

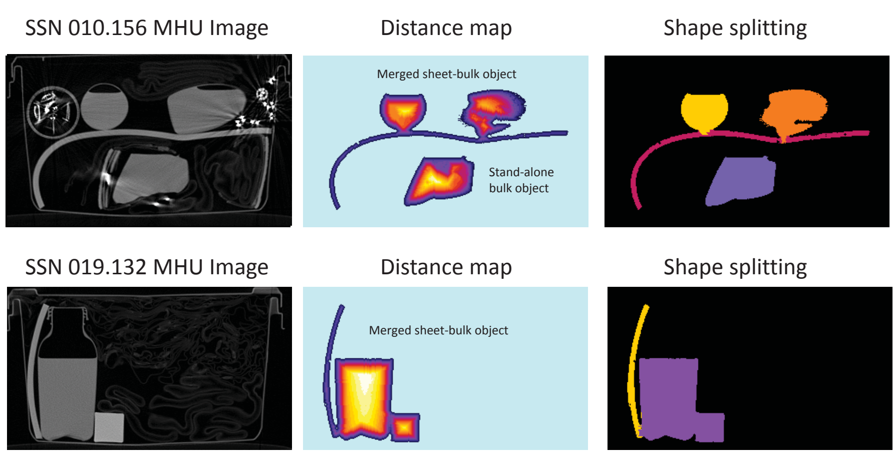
Object Splitting 1: Graph Based

- Label image graph model
  - Vertices: object pixels
  - Edges: pixel connectivity
- BFS: Breadth-First Search
  - a) Compute distance from surface to object core
  - b) Grow multiple objects back from select core regions
- DFS: Depth-First Search
  - c) Detect sheet by finding equidistant surfaces from prespecified core depths



11/6/2014 Jens Gregor 7

Separating Sheet and Bulk Objects

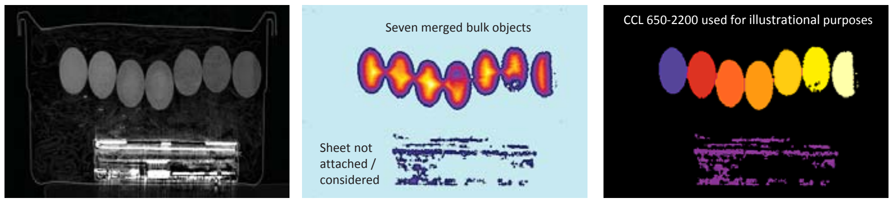


Top: Sheet is separated from small bulk objects attached to it
Bottom: Sheet is separated from large bulk object attached to it

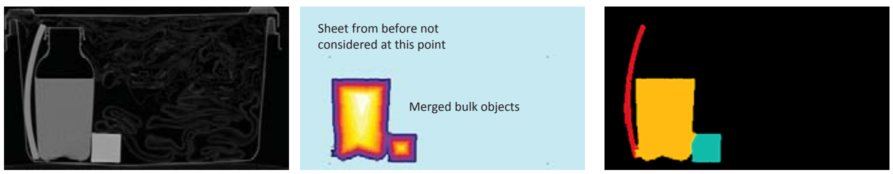
11/6/2014 Jens Gregor 8

Separating Touching Bulk Objects

SSN 012.YZ346 MHU Image | Distance map | Shape splitting



SSN 019.132 MHU Image | Distance map | Shape splitting

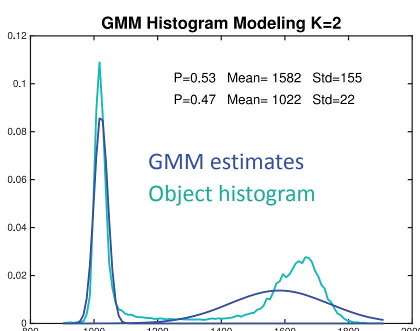


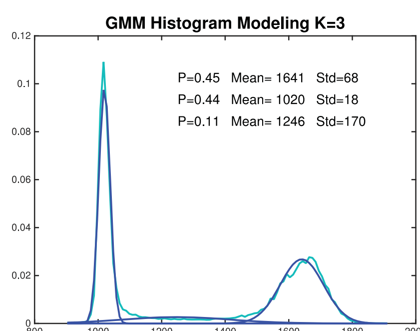
Top: Pseudo-targets (powder) are separated from one another
Bottom: Two bulk objects are separated from one another

11/6/2014 Jens Gregor 9

Object Splitting 2: GMM Based

Gaussian mixture model: statistical parameter estimation (ML-EM)

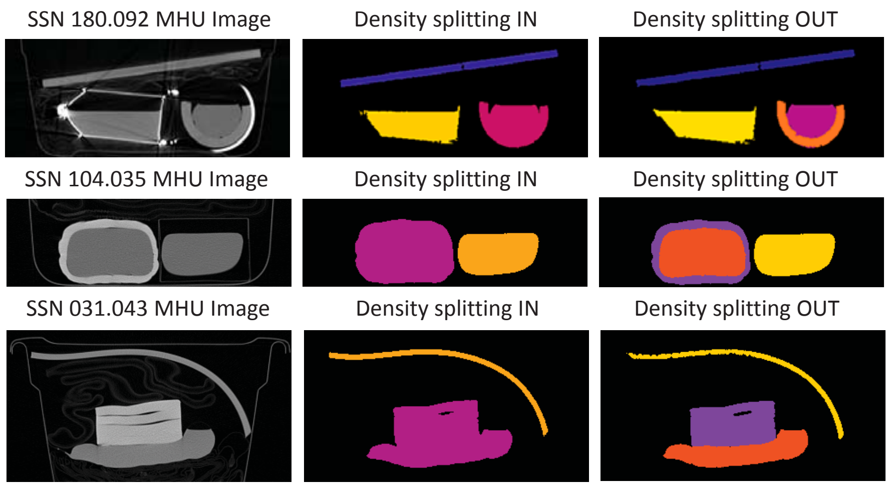




Using two Gaussians yields separation. Using three Gaussians allows gap between peaks to be modeled. Computational cost managed by sampling.
Separation is more involved than Bayesian min prob(error) thresholding.

11/6/2014 Jens Gregor 10

Separating Different Materials



Combination of shape and density splitting recovers wide variety of objects

11/6/2014 Jens Gregor 11

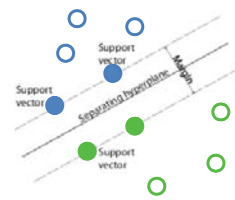
Feature Extraction/Description

- Features extracted during object formation
  - Num. voxels
  - Mass
  - Density mean
- Features extracted prior to classification
  - Object type: sheet or bulk
    - Number of “deep core” voxels relative to total number of voxels: sheet = 0-5, bulk object=5-35

11/6/2014 Jens Gregor 12

SVM Classifier Primer

- Maximum margin linear separation of feature vectors for two classes
- Kernel function, eg Gaussian RBF, maps problem to non-linear space
- Slack variables yield regularization allowing for imperfect separation



http:://mathworks.com

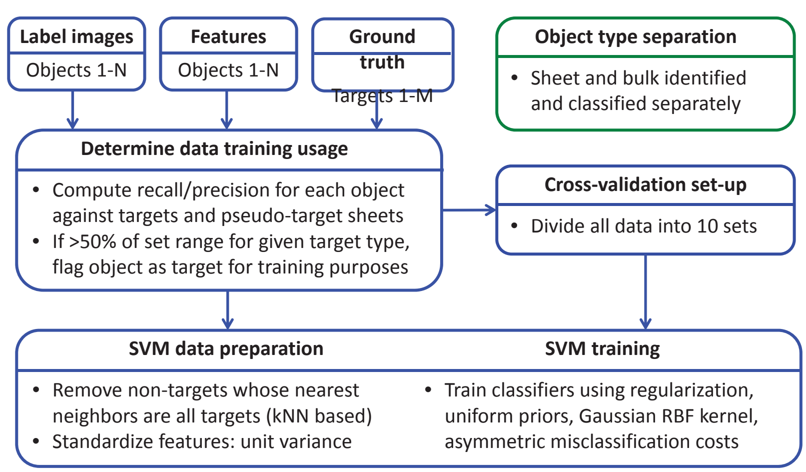
Constrained quadratic programming solved using Lagrange dual problem formulation of the primal problem given by

$$\operatorname{argmin} \tfrac{1}{2}\|w\|^2 + C\sum s_i \quad \text{s.t.} \quad y_i\,(w^T\Phi(x_i) - b) \geq 1 - s_i;\ s_i \geq 0$$

V. Vapnik, "The Nature of Statistical Learning Theory", Springer, 1995
J. Gregor & Z. Liu, "Regularized SVM Param. Est.", IEEE Intl. Conf. Data Mining, 2004

11/6/2014 Jens Gregor 13

ATR Training Outline



11/6/2014 Jens Gregor 14

ATR Data Preparation

- Training targets identified as objects for which recall, precision > 0.50 min limits for ground truth
  - Using lower threshold moves decision boundary away from true target which may improve classification
- Non-targets surrounded by targets are discarded from consideration using 3-NN preclassification
  - Reduces confusion between targets and non-targets which leads to SVM separation in favor of target
  - Stricter rule (5-NN) applied to pseudo target sheets defined here as mass < 125 g, 1100 < density < 1300

11/6/2014 Jens Gregor 15

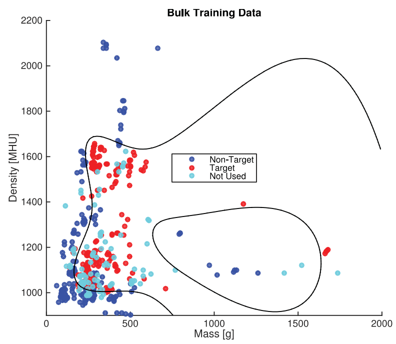
SVM Training

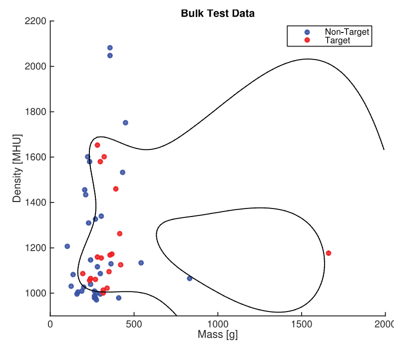
- Automated separation of objects as sheet or bulk followed by training of separate SVM classifiers
- SVM classifier configuration
  - Uniform priors: removes empirical bias
  - Gaussian RBF kernel: data similarity model
  - Regularization: soft margin classification
  - Misclassification costs: target errors set to be 2x more costly than non-target errors

11/6/2014 Jens Gregor 16

Example Scatter Plots for Bulk

SVM decision boundaries shown for one training data set and one test data set using mass and density features. Notice discarded non-target training samples.



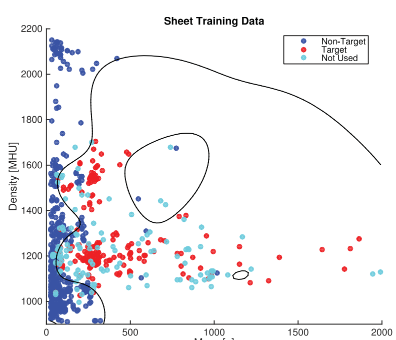


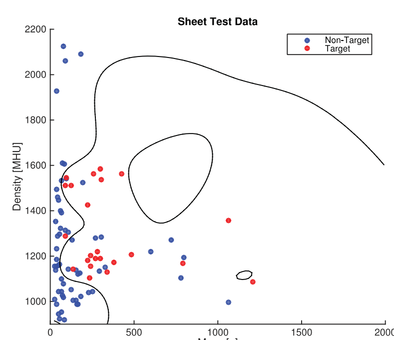
NOTE: This is an illustrative example only. Actual classifier uses more features and different parameter settings.

11/6/2014 Jens Gregor 17

Example Scatter Plots for Sheets

SVM decision boundaries shown for one training data set and one test data set using mass and density features. Notice discarded non-target training samples.





NOTE: This is an illustrative example only. Actual classifier uses more features and different parameter settings.

11/6/2014 Jens Gregor 18

PD/PFA results

Num Non-targets	Num FAs	PFA [%]
1371	274	20

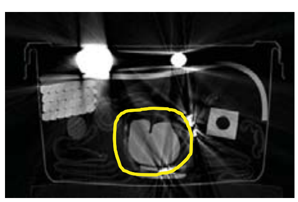
Target Type	Target Subtype	Level of Difficulty	Num Targets	Num Detected	PD [%]
● **Target**	**All**	**All**	**407**	**359**	**88.2**
Target	Clay	All	111	95	85.6
Target	Rubber	All	158	148	93.7
Target	Saline	All	138	116	84.1
Target	Bulk	All	270	229	84.8
● **Target**	**Sheet**	**All**	**137**	**130**	**94.9**
Target	All	Low	77	68	88.3
Target	Clay	Low	29	26	89.7
Target	Rubber	Low	22	21	95.5
Target	Saline	Low	26	21	80.8
Target	Bulk	Low	56	47	83.9
Target	Sheet	Low	21	21	100
Target	All	High	317	279	88
Target	Clay	High	82	69	84.1
Target	Rubber	High	125	117	93.6
Target	Saline	High	110	93	84.5
Target	Bulk	High	201	170	84.6
Target	Sheet	High	116	109	94
● **Pseudo Target**	**Sheet**	**High**	**10**	**7**	**70**

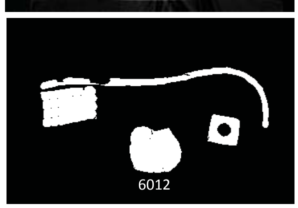
11/6/2014 Jens Gregor 19

Case 1: SSN 013 Slice 105

Target 6012: 10% saline solution in breast milk bag; Mass 285 g

Detected Recall=78%, Precision=95%; Mass=205 g





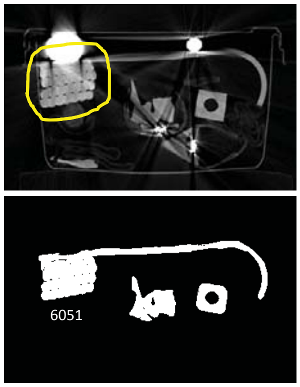
Missed object touches sheet elsewhere.
Objects both made of rubber.

11/6/2014 Jens Gregor 20

Case 2: SSN 013 Slice 128

Target 6051: Clay object; Mass 286 g

Detected Recall=81%, Precision=89%; Mass=345 g



Same missed detection as for Case 1

11/6/2014 Jens Gregor 21

Case 3: SSN 035 Slice 049

Target 6150: Clay object; Mass 290 g

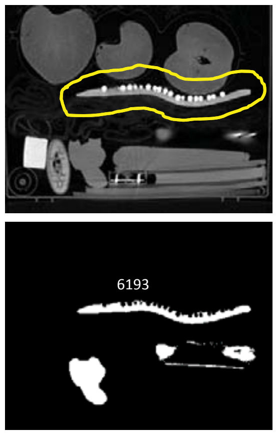
Detected Recall =70%, Precision=80%; Mass=304 g

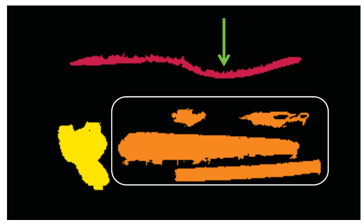
11/6/2014 Jens Gregor 22

Case 4: SSN 193 Slice 198

Target 6193: Clay object; Mass 410 g

Detected Recall=69%, Precision=99%; Mass=298 g





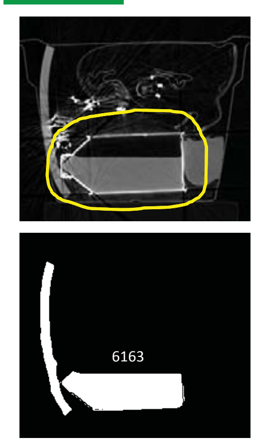
Missed detection/False alarm

11/6/2014 Jens Gregor 23

Case 5: SSN 063 Slice 045

Target 6163: 5% saline solution in tin bottle; Mass 274 g

Detected Recall=81%, Precision=89%; Mass=249 g

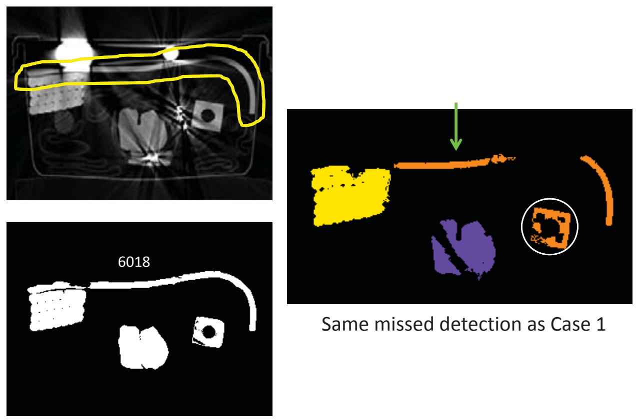


11/6/2014 Jens Gregor 24

Case 6: SSN 013 Slice 111

Target 6018: 1/4” rubber sheet; Mass 685 g

Detected Recall=41%, Precision=55%; Mass=603 g



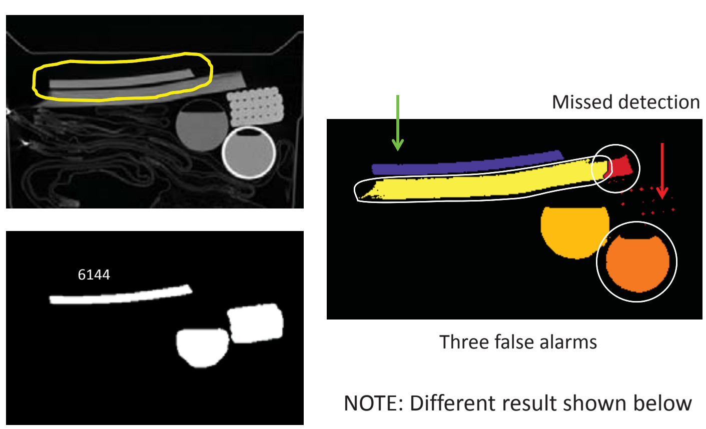
Same missed detection as Case 1

11/6/2014 Jens Gregor 25

Case 7: SSN 033 Slice 046

Target 6144: 3/8” rubber sheet; Mass 345 g

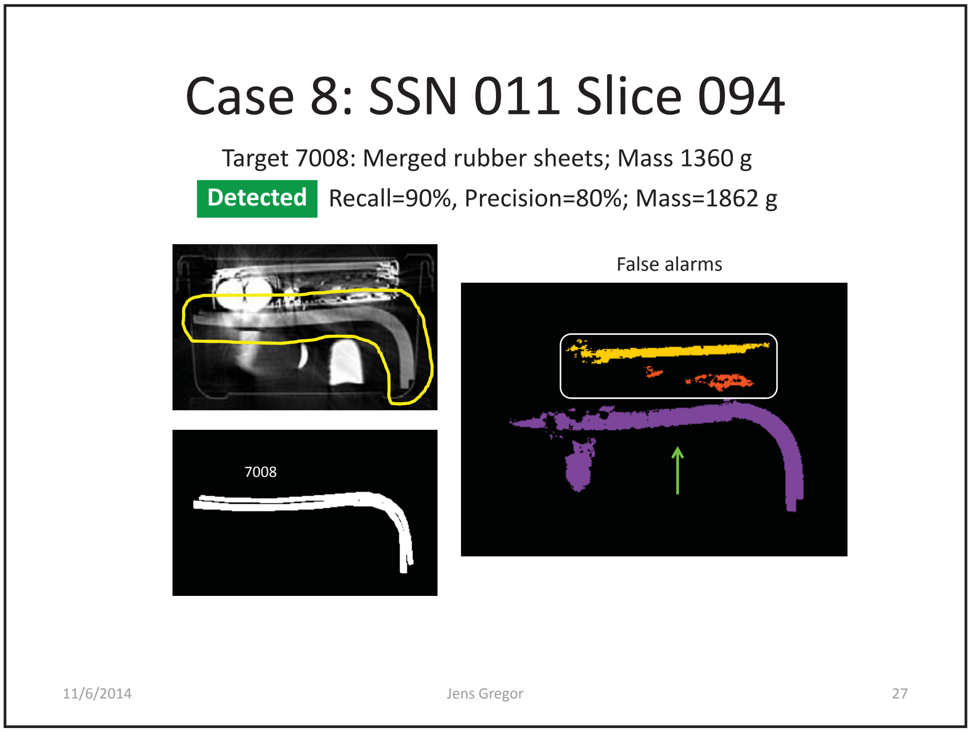
Detected Recall=27%, Precision=96%; Mass=95 g

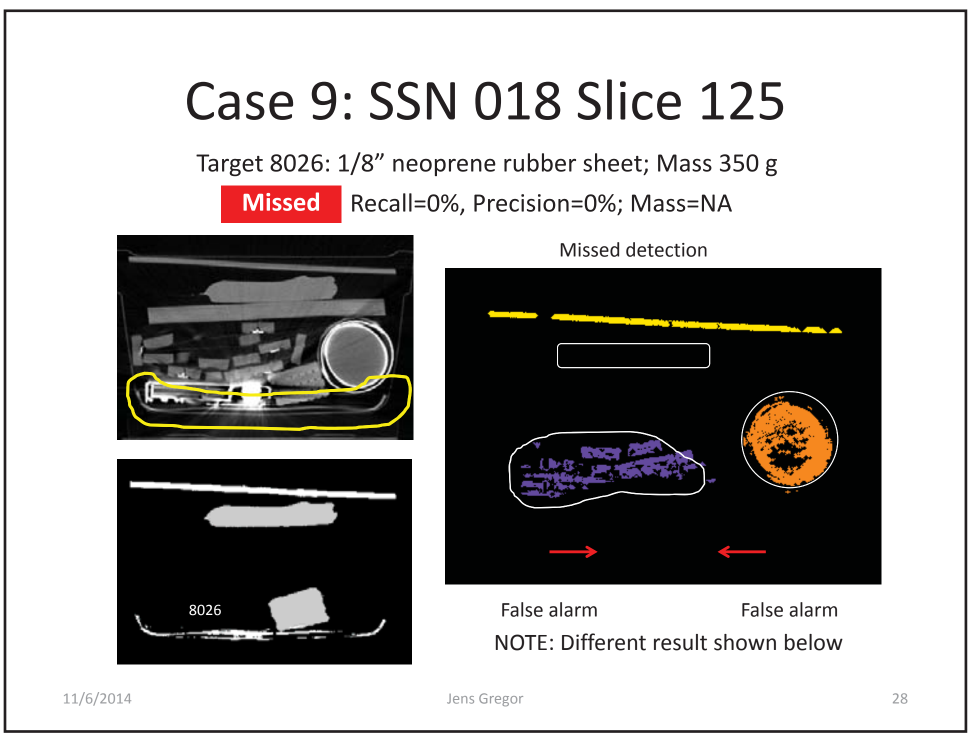


Three false alarms

NOTE: Different result shown below

11/6/2014 Jens Gregor 26

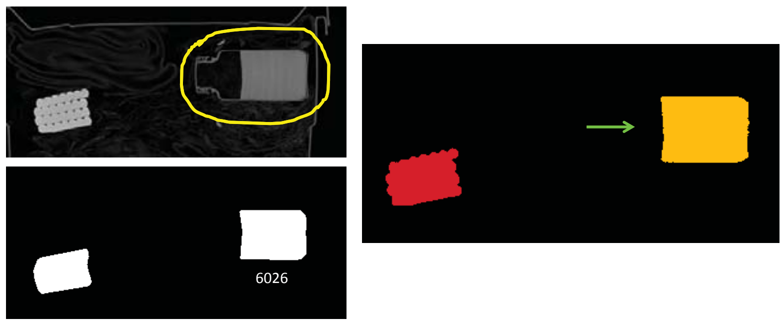

Case 8: SSN 011 Slice 094
Target 7008: Merged rubber sheets; Mass 1360 g
Detected
Recall=90%, Precision=80%; Mass=1862 g
False alarms
7008
11/6/2014
Jens Gregor
27


Case 9: SSN 018 Slice 125
Target 8026: 1/8” neoprene rubber sheet; Mass 350 g
Missed
Recall=0%, Precision=0%; Mass=NA
Missed detection
8026
False alarm
False alarm
NOTE: Different result shown below
11/6/2014
Jens Gregor
28

Case 10: SSN 012 Slice 105

Target 6026: Powder pseudo target; Mass 277 g

Detected Recall=97%, Precision=98%; Mass=206 g



11/6/2014 Jens Gregor 29

Strengths and Weaknesses

✓ Advantages associated with proposed approach
  - Good segmentation facilitates simple classifier
  - Simple classifier with easy to understand behavior

- Limitations wrt target type, mass, density, etc
  – SVM classifier will tune itself to the training data
- General reasons for missed targets/false alarms
  – Low recall, precision due to imperfect segmentation
  – Poor separation of target and non-target features

11/6/2014 Jens Gregor 30

Possible ATR Improvements

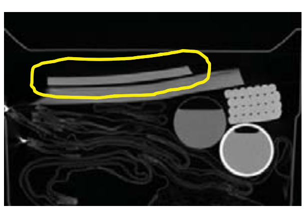
- Segmentation
  - Add metal artifact reduction and handling
  - Make shape based splitting density aware and likewise make density splitting shape aware
- Classification
  - Identify more/more discriminatory features
  - SVM: Applying weighting to the data samples
  - SVM: Apply optimization to parameter settings

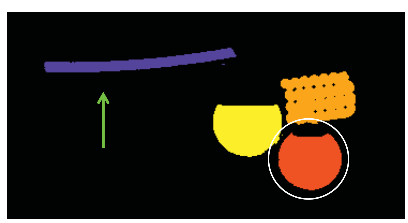
11/6/2014 Jens Gregor 31

Alt. Case 7: SSN 033 Slice 046

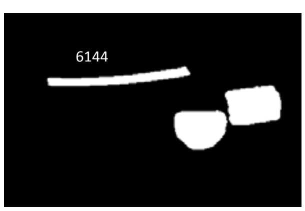
Target 6144: 3/8” rubber sheet; Mass 345 g

Detected Recall=27%, Precision=96%; Mass=95 g





False alarm



Modified classifier (more features, less strict SVM training) captures missed clay object (orange) and reduces false alarms from three to just one.

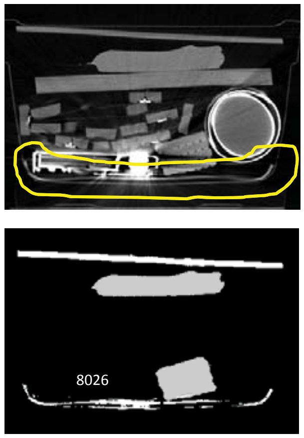
OLD:20/88/95/70 NEW:15/81/87/90

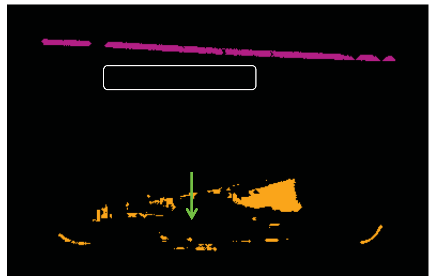
11/6/2014 Jens Gregor 32

Alt. Case 9: SSN 018 Slice 125

Target 8026: 1/8" neoprene rubber sheet; Mass 350 g

Detected Recall=29%, Precision=11%; Mass=1172 g





Modified classifier (more features, less strict SVM training, data recall-precision weighting) captures missed pseudo target sheet, eliminates false alarms.

OLD: 20/88/95/70 NEW: 22/86/90/100

11/6/2014 Jens Gregor 33

Comments on Data and Tools

- Many thanks for the hard work several people put into generating the data sets as well as the automated tools for testing ATR performance!!
- Having same or similar material be both target and non-target increased the level of difficulty
  - Rubber sheets (target) vs rubber soles (non-target)
  - Low saline solutions (targets) vs water (non-target)
- Pseudo targets well outside range of targets

11/6/2014 Jens Gregor 34

Lessons Learned

- ATR design and implementation is very difficult
  - Context dependent data appearance (shape, value)
  - Target characteristics very similar to non-targets
- Image quality does not predict ATR performance
  - Good segmentation needed but better segmentation does not necessarily guarantee improved PD/PFA
- Many moving parts – all must be perfectly tuned
  - Knowing when (not) to execute task is important
  - What improves one case may work against another

11/6/2014 Jens Gregor 35

Summary

New ATR developments have been presented:

1. Shape based segmentation using graph traversal algorithms (breadth-first and depth-first search)
2. Gaussian mixture model for material type based segmentation (density histogram modeling)
3. Regularized SVM classifier (k-NN data editing)

Future work on PFA reduction needed including:

1. Improvement on image quality (metal artifacts)
2. Use of more discriminatory density features

11/6/2014 Jens Gregor 36

11.5.2.5 ATR Development – Zhang

"ATR Development: University of Wisconsin-Milwaukee"

ATR Development
University of Wisconsin-Milwaukee

Jun Zhang

Laura Drake

Hongquan Zuo

1

Institution and Researchers

- Institution: University of Wisconsin-Milwaukee
  - 2nd largest state school in Wisconsin (1st: Madison)
- Jun Zhang, Professor
  - PhD,EE, Rensselaer Polytechnic Institute
  - Area of interest: signal/image processing
  - Work done for ALERT
    - This work: funded by ALERT
    - Previous work: reconstruction, unfunded
- Laura Drake, Researcher
  - PhD, EE, Northwestern University
  - Area of interesting: signal/image processing
  - Prior work: Space Applications Co., CA
- Hongquan Zuo, PhD student
  - Area of interest: image processing
- We also thank Chris Kallas and Yingying Gu for their help

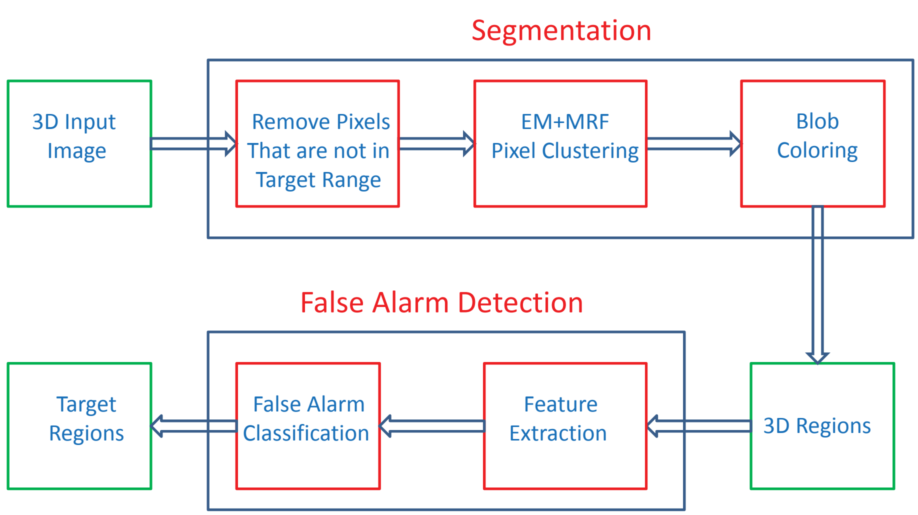
2

PD/PFA Results: 89%PD, 10%PFA, 80%PD for PT

Target Type	Target Subtype or Form	Level of Difficulty	Num Targets	Num Detected	PD [%]
Target	All	All	407	363	89.2
Target	Clay	All	111	105	94.6
Target	Rubber	All	158	142	89.9
Target	Saline	All	138	116	84.1
Target	Bulk	All	270	237	87.8
Target	Sheet	All	137	126	92
Target	All	Low	77	73	94.8
Target	Clay	Low	29	29	100
Target	Rubber	Low	22	22	100
Target	Saline	Low	26	22	84.6
Target	Bulk	Low	56	52	92.9
Target	Sheet	Low	21	21	100
Target	All	High	317	278	87.7
Target	Clay	High	82	76	92.7
Target	Rubber	High	125	110	88
Target	Saline	High	110	92	83.6
Target	Bulk	High	201	173	86.1
Target	Sheet	High	116	105	90.5
Pseudo-target	Sheet	High	10	8	80
Num Non-targets	Num FAs	PFA [%]			
1370	133	9.7			

3

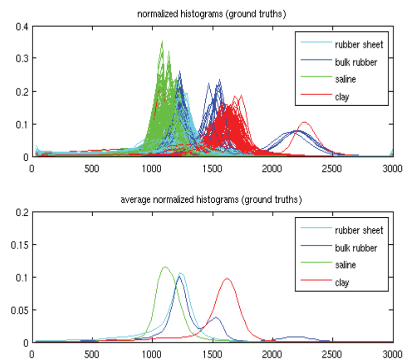
ATR: High-Level Description



- Pixel clustering/classification with spatial smoothness constrains
- False alarm Classification using histogram and gradient features

Technical Description: Starting with Observations of Histograms

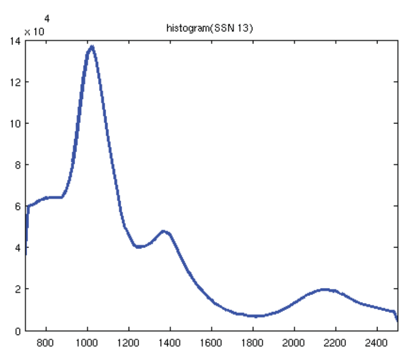
Histograms, i.e., *intensity distributions*, of ground truth targets and their averages



- Most target pixels have intensities in [a, b] (e.g., a=700, b=2500);
- Target-pixel histograms, i.e., intensity distributions, are Gaussian-like (or , for bulk rubber, Gaussian mixture like);
- Some of them overlap (e.g., for Saline and rubber)
- The ground truth targets contain artifacts, which increase the variances of their intensity distributions/histograms

Observations (continued)

The histogram, i.e., the intensity distribution, of an Image (SSN 13)



- The histogram, i.e., intensity distribution, contains multiple “peaks” or “hills”
- Each peak is Gaussian-like
- Each peak is generally made up by pixels from a single object or, sometimes, from more than one object with similar intensity distributions

6

Our Approach to Segmentation The Basic Idea

- What is segmentation?
  - Breaking a 3D CT image into regions that, hopefully, correspond to targets and other objects
- Based on observations in the previous two slides, segmentation can be achieved by
  - Identifying the "histogram peaks" of the image
  - Assigning/classifying each pixel of the image to one of these peaks
  - Grouping neighboring pixels into the same region if they belong to the same peak
  - For best results, segmentation need to iterate between peak identification and pixel classification

7

Segmentation – "high-level" steps

- Our approach:
  - Step 1: pixels with intensities outside the "target range" (e.g., outside [700, 2500]) are labeled as background
  - Step 2a: the rest of the pixels ("foreground") are used to identify "histogram peaks," i.e., intensity classes
  - Step 2b: the intensity classes are used to classify each pixel
  - Step 3: neighboring pixels in the same class are grouped by blob coloring (i.e., CCL) into regions
- Our presentation will be focused on Step 2, since Steps 1 and 3 are straightforward

8

Step 2 Pixel Classification
-- A "High-Level" Description

- The "histogram peaks" of an image can be modeled by Gaussian distributions, each with a mean and a variance
- The image pixels are used to estimate these means and variances using an algorithm known as the EM (expectation-maximization)
- The EM algorithm also provides a classification for each pixel, i.e., to which Gaussian peak it belongs to
- Finally, the MRF is incorporated into the EM to encourage spatially smooth classification – neighboring pixels are more likely to belong to the same class (i.e., histogram peak)

9

Pixel-Classification
-- A Mathematical Description

- Some notation:
  - y = {y_i} = image pixels, i denotes pixel location
  - theta = parameters the "peaks" of the histogram of the image, i.e., the mean and variance of the Gaussian distributions
  - z = {z_i} = class labels, z_i = k indicates pixel y_i is assigned to the kth peak of the histogram of the image
  - theta = parameter vector for the "histogram peaks", i.e., the mean and variance of the Gaussian distributions
- The segmentation problem:
  - Finding z and theta from y
  - How? By maximizing the likelihood function on the right hand side:

$$\hat{\theta} = \arg\max_{\theta} \log p(y|\theta) = \arg\max_{\theta} \log \left\{ \sum_{z} p(y, z|\theta) \right\}$$

  - The density on the right hand side is a product of Gaussian mixtures
  - How to solve the above optimization problem? Use the EM algorithm
  - The EM algorithm also allows to find z, the segmentation

10

The EM Algorithm

- The EM algorithm iterates between two steps
  - The E-step: compute the class-probability of each pixel

$$r_{i,k} = P[z_i = k] = Cp(y_i | z_i = k, m_k, \sigma_k)\pi_k$$

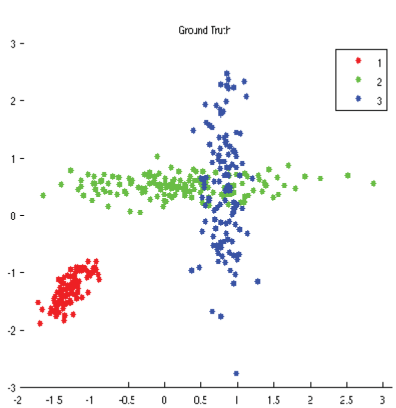
  - The M-step: compute the mean, variance, and proportion of each class

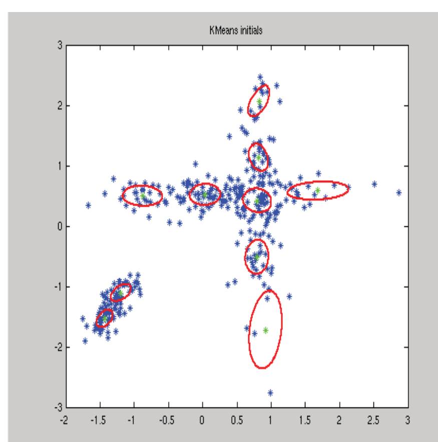
$$m_k = \frac{\sum r_{i,k} y_i}{\sum r_{i,k}}, \quad \sigma_k^2 = \frac{\sum r_{i,k}(y_i - m_k)^2}{\sum r_{i,k}}, \quad \pi_k = \frac{\sum r_{i,k}}{n}$$

  - Final segmentation: $z_i = \arg\max_k r_{i,k}$
  - Advantage: allows class/cluster overlap
  - Problem: noisy segmentation

11

An Example in Video (GIF files): the EM algorithm estimates the 2D Gaussian clusters of a mixture





3-cluster mixture data
The ground truth

EM results: each estimated Gaussian cluster is represented by an ellipse

12

Adding the MRF

- The EM algorithm used alone tends to produce "noisy" segmentations
- Solution: use a "Markov random field (MRF) prior" to make segmentation smooth.
- MRF prior on segmentation

$$p(\{z_i\}) = \frac{e^{-\beta U(\{z_i\})}}{Z}$$

$$U(\{z_i\}) = \sum \left\{ V_1(z_i) + \sum_{j \in N_i} V_2(z_i, z_j) \right\}$$

  – The 2nd term in the sum encourages smooth segmentations ({z}), where neighboring pixels tend to be assigned to the same intensity class

13

After Combining EM with MRF

- The E-step of the now EM is changed to

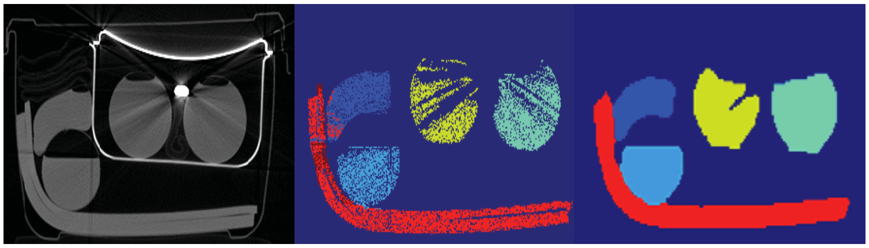
$$r_{i,k} = P[z_i = k] = C \sum_{\{z_j\} - z_i} p(y_i | z_i = k, m_k, \sigma_k) p(\{z_j\})$$

- No change to the M-step
- Problem: the number of terms in the sum for the new E-step is too large
- Solution: use approximation (e.g., the mean field theory [MFT])

14

Example: MRF-EM Segmentation

- Input image: SSN 007



- Images: input image, EM only segmentation, MRF-EM segmentation

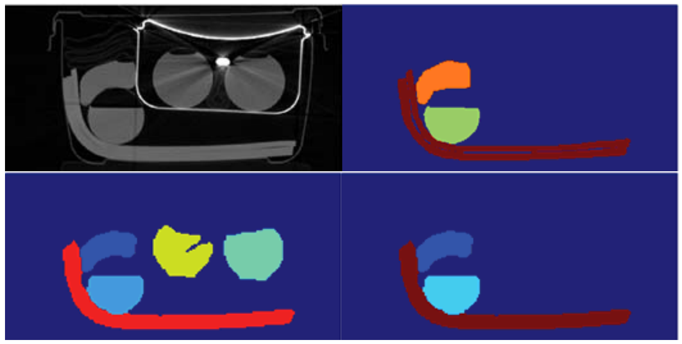
15

Segmentation Algorithm Recap

- Input: a 3D image
- Output: 3D regions
- Steps
  - Step 1: Put pixels outside target range as background
  - Step 2: perform MRF-EM pixel classification on foreground pixels
  - Step 3: use blob coloring (i.e., CCL) to group spatially adjacent same-class pixels into regions
- An example: next page

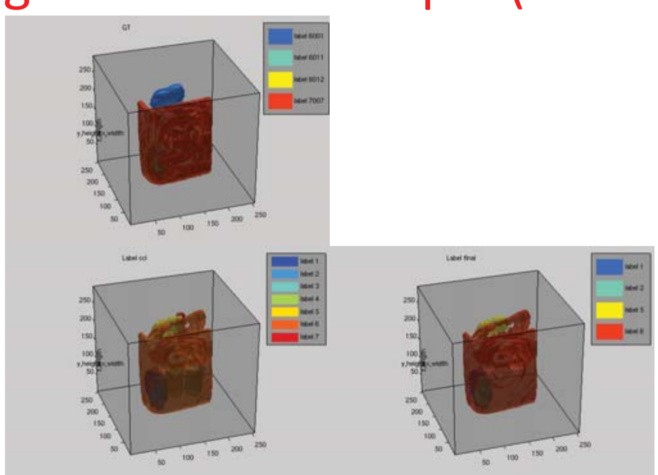
16

Example: Segmentation Results



- •Top row: original image; ground truth
- •Bottom row: MRF-EM segmentation + CCL; after FA detection (to be described later)
- •Observations:
  - •Pixels with similar intensities are grouped into regions
  - •Resistant to some metal artifacts

Segmentation Example (continued)



- •Top: ground truth
- •Bottom row: MRF-EM segmentation + CCL, FA Rejection result
- •Observation: targets detected; merge with shoe soles due to intensity similarity

Other Issues Related to Segmentation

- Splitting and merging
  - Addressed implicitly through parameters of the segmentation algorithm, such as the number of classes and beta (the amount of smoothing)
- Detecting and correcting CT artifacts: no
- Shape: not used
- Addressed bulks and sheets differently: no
- Detection of CT artifacts: no
- Special processing for pseudo target sheets: none

19

False Alarm Detection after Segmentation

- Segmentation Results and PD/PFA:
  - Segmentation produces two type of regions: targets and false alarms
  - It is difficult to obtain 100% PD
- Solution:
  - First, make PD as large as possible in segmentation
  - Then, reduce PFA by detecting false alarm regions
- False alarm detection
  - Input: regions from segmentation and their features
  - Output: false alarm classification
  - Final outcome: segmented regions labeled as false alarms are rejected

20

False Alarm Detection/Classification

- FA detection/classification
  - Region features: Histogram, average gradient, mean, variance, histogram mode, and region size
  - Classifier: SVM (support vector machines) trained on 50% of randomly selected segmented region/feature data
  - FA classifier pd/pfa and overall ATR PD/PFA

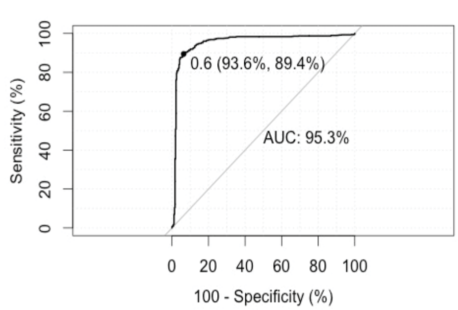
$$PD_{ATR} = PD_{seg}(1-pfa), \quad PFA_{ATR} = PFA_{seg}(1-pd)$$

  - Select FA classifier operating point (pfa,pd) on its ROC curve to achieve desired final ATR PD/PFA

21

False Alarm Detection Example

- FA detector ROC curve
  - Selected operating point: pfa=3%, pd=80%

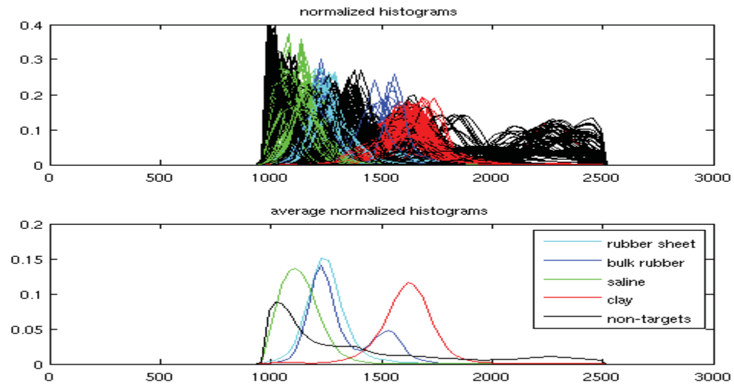


- Overall ATR PD/PFA
  - Before FA detection: 89%PD, 62%PFA
  - After: 86%PD, 12%PFA

22

One Step Further: Collective FA Classification Starting with Observations

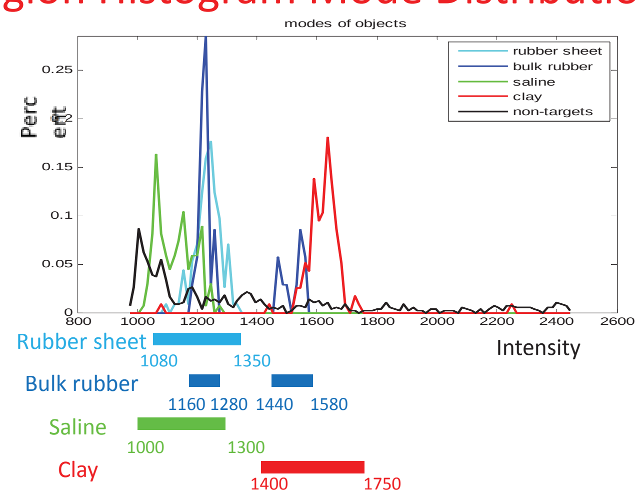
Histograms of segmented regions



Observation: histogram for different targets are in different intervals (overlaps do occur); histograms for FA regions are in all intervals

23

More Observations: Region Histogram Mode Distributions



Observation: histogram modes of different targets are often in different intervals (although overlaps do occur sometimes); for FA regions, the histogram mode spreads out in all regions

24

Now: Collective FA Classification

- Method:
  - Divide the foreground intensity range into several (some overlapping) target intervals, one for each target type, e.g., [1080, 1350] for rubber sheet (see previous slide)
  - Have a FA classifier for each interval, 5 all together (see previous slide)
  - For each segmented region, send it to a classifier according to its histogram mode, e.g., if its mode is in [1080, 1350], send it to that interval classifier
  - Since the intervals can overlap, a region may be classified by more than one classifier; in this case, it will be a "target" if at least one classifier says it is
  - How is the classifier trained? See next slide

25

Collective FA Classification (Continued)

- Training the collective classifiers:
  - The training data contain randomly selected segmented regions, some target and some false alarm (see Slide 21)
  - Divide the training data set into 5 subsets corresponding to the 5 target intervals
  - A (training region) is in the *k*th subset if its histogram mode falls into the *k*th interval
  - Use the *k*th subset to train the *k*th FA classifier
- Some example results:
  - After segmentation: 90%PD, 70% PFA
  - After FA classification: 89%PD, 10% PFA
  - Another example: (92%PD, 42%PFA) after segmentation, classification, (90%PD, 9%PFA) after FA classification

26

Relation to Prior Art and Innovation

- Previous approach in segmentation in general
  - Region growing, pixel classification/clustering, curve evolution, graph cuts
- Prior patents
  - Region growing, special processing: bulks and sheets, post segmentation classification for false alarm rejection
- What's new about our Approach
  - Segmentation: the MRF-EM is based on own work
  - False alarm classification: our collective FA classification are not seen in previous literature
  - Both our segmentation and false alarm classification methods are not used in prior patents

27

Difference from other 4 ATRs

- For segmentation: we used pixel classification
  - Segmented regions are made up with pixels of different intensity classes
  - MRF prior is used to get smooth segmentations and to deal with some of the metal artifacts
- For false alarm reduction: we used collective classification
  - The multiple classifiers are based the intervals of the histogram modes of the segmented regions
  - Each region can be classified by multiple classifiers
- Our technique has only two components: region segmentation and region (FA) classification
- The segmentation part is unsupervised (no training)

28

ATR Training

- Segmentation algorithm
  - Unsupervised: no training
- False alarm classification
  - Took half of the regions data randomly as training data
- For new type of targets
  - Segmentation algorithm requires no change
  - FA classification may or may not need to be retrained, depending on whether the intensities of the objects changes significantly

29

Some Other Issues

- How is shape used?
  - We did not use shape
- How would ground truth for non-targets improve our ATR?
  - Not needed for our segmentation
  - Might be useful for our false alarm detection but not critical
- How can PD and PFA be improved without over training?
  - Since most of the targets missed by segmentation is due to merge, metal artifact reduction and 3D edge detection might improve on PD; no training is needed
  - This can also reduce PFA, no additional training is needed

30

PD/PFA Results: 89%PD, 10%PFA, 80%PD for PT

Target Type	Target Subtype or Form	Level of Difficulty	Num Targets	Num Detected	PD [%]
Target	All	All	407	363	89.2
Target	Clay	All	111	105	94.6
Target	Rubber	All	158	142	89.9
Target	Saline	All	138	116	84.1
Target	Bulk	All	270	237	87.8
Target	Sheet	All	137	126	92
Target	All	Low	77	73	94.8
Target	Clay	Low	29	29	100
Target	Rubber	Low	22	22	100
Target	Saline	Low	26	22	84.6
Target	Bulk	Low	56	52	92.9
Target	Sheet	Low	21	21	100
Target	All	High	317	278	87.7
Target	Clay	High	82	76	92.7
Target	Rubber	High	125	110	88
Target	Saline	High	110	92	83.6
Target	Bulk	High	201	173	86.1
Target	Sheet	High	116	105	90.5
Pseudo-target	Sheet	High	10	8	80
Num Non-targets	Num FAs	PFA [%]			
1370	133	9.7			

31

10 Cases

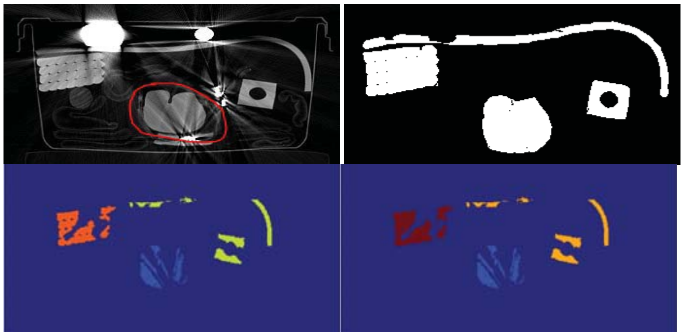
- Summary

Case Number	Target Detected?
1	Yes
2	Yes
3	No
4	Yes
5	Yes
6	Yes
7	Yes
8	Yes
9	No
10	Yes

- Details: next page

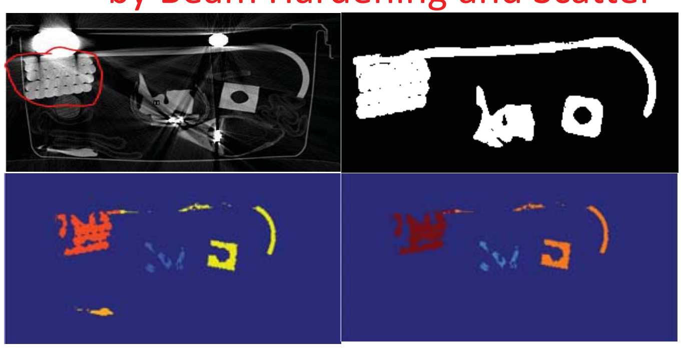
32

Case 1: SSN 13, Bulk with Bad Streaks



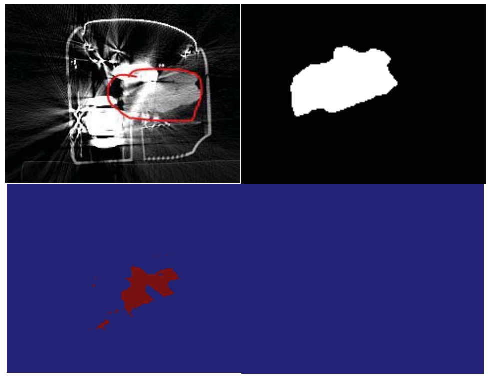
- First row: CT image and ground truth (GT)
- Second row: segmentation and after FA detection
- Target Detected? Yes.
- Detected by segmentation (with P=0.95, R=0.51; low recall due to detection window) and kept by FA detection

Case 2: SSN 13, Bulk with Bad Shading Caused by Beam Hardening and Scatter



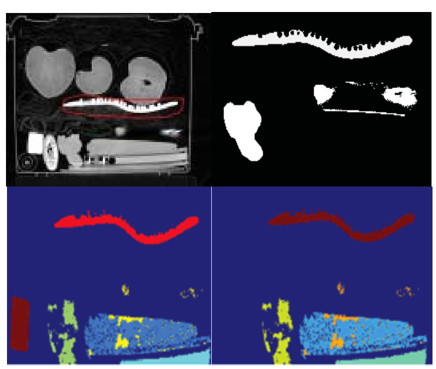
- First row: CT image and ground truth (GT)
- Second row: segmentation and after FA detection
- Target Detected? Yes.
- Detected by segmentation (with P=0.90, R=0.59; low recall due to detection window) and kept by FA detection

Case 3: SSN 35, Bulk inside Electronics



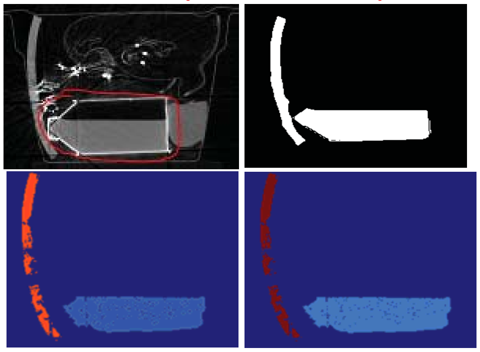
- First row: CT image and ground truth (GT)
- Second row: segmentation and after FA detection
- Target Detected? No.
- It was there after segmentation (with P=0.92, R=0.45; low recall due to metal artifacts). Our FA classifier labeled it as an FA and removed it.

Case 4: SSN 193, Bulk with Texture



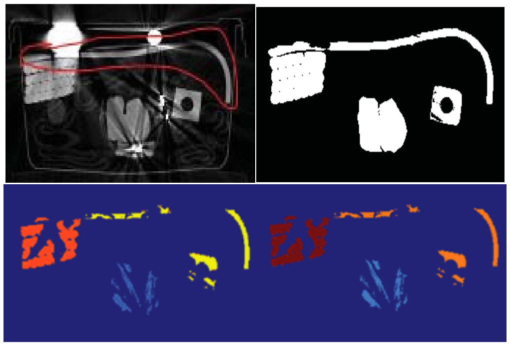
- First row: CT image and ground truth (GT)
- Second row: segmentation and after FA detection
- Target Detected? Yes.
- Detected by segmentation (with P=0.99, R=0.66) and kept by FA detection.

Case 5: SSN 63, Bulk with density close to water (~5% saline)



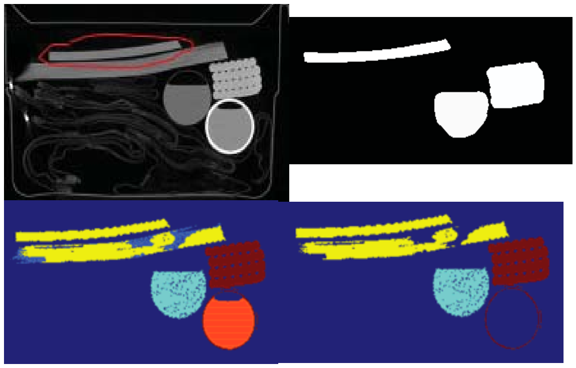
- First row: CT image and ground truth (GT)
- Second row: segmentation and after FA detection
- Target Detected? Yes.
- Detected by segmentation (with P=1.00, R=0.61) and kept by our FA detector.

Case 6: SSN 13, Sheet with bad streaks caused by metal, beam hardening and scatter



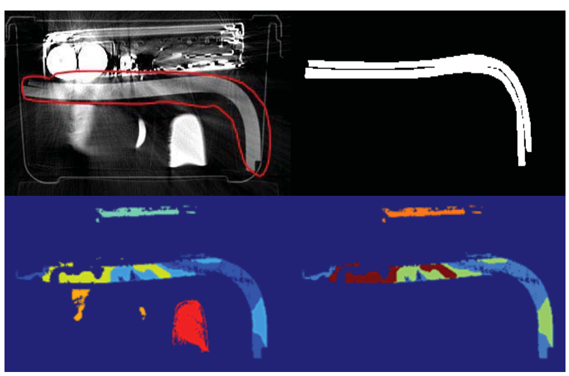
- First row: CT image and ground truth (GT)
- Second row: segmentation and after FA detection
- Target Detected? Yes.
- Detected by segmentation (with P=0.49, R=0.26) and kept by our FA detector.

Case 7: SSN 33, Sheet laying on top of another flat object



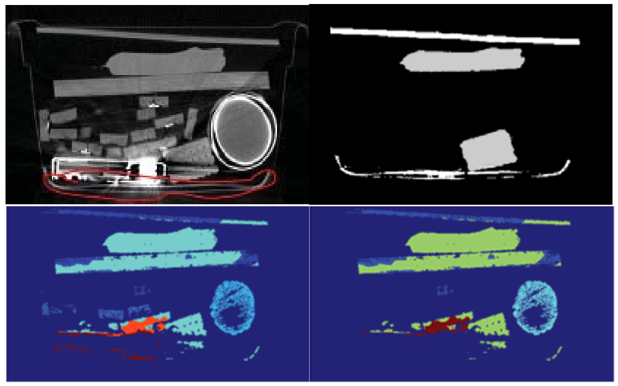
- First row: CT image and ground truth (GT)
- Second row: segmentation and after FA detection
- Target Detected? Yes.
- Detected by segmentation (with P=0.28, R=0.74) and kept by our FA detector.

Case 8: Object with lots of photon starvation



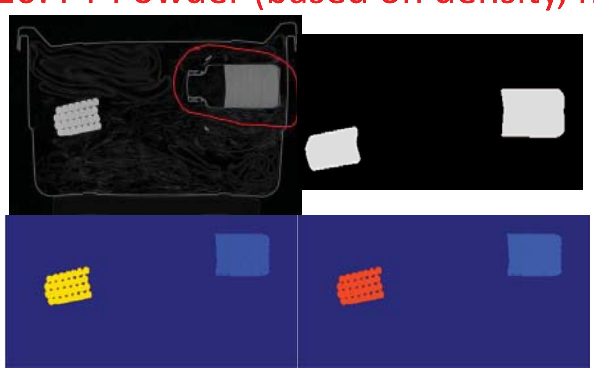
- First row: CT image and ground truth (GT)
- Second row: segmentation and after FA detection
- Target Detected? Yes.
- Detected by segmentation (with P=0.86, R=0.43) and kept by our FA detector.

Case 9: PT sheet based on thickness



- First row: CT image and ground truth (GT)
- Second row: segmentation and after FA detection
- Pseudo-targets Detected? No.
- Not detected by segmentation (with P=0.10, R=0.26), low-recall due to metal artifacts.

Case 10: PT Powder (based on density, not mass)



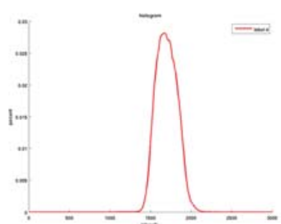
- First row: CT image and ground truth (GT)
- Second row: segmentation and after FA detection
- Pseudo-target Detected? Yes.
- Detected by segmentation (with P=0.99, R=0.93) and kept by our FA detector.

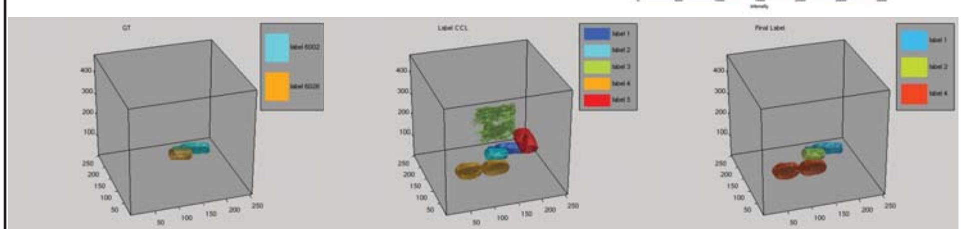
Two False Alarm Cases

- Next page

43

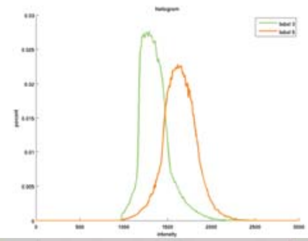
FA in SSN 008

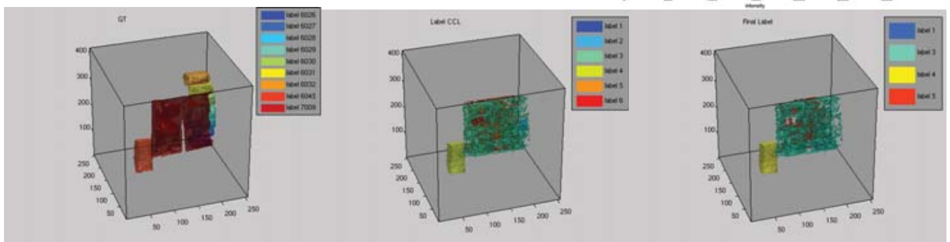




•Top row: CT image slice for SSN 88, histogram of the FA
•Bottom row: ground truth, ATR output, two FA objects
•FA objects: orange
•Observations: the FA objects have similar intensity as clay targets; our FA detection selected a pfa (to preserve high ATR PD) hence cannot detect all FAs

FAs in SSN 012





•Top row: image slice for SSN 012, histogram of the FAs
•Bottom row: ground truth, ATR output, two FA objects
•FA objects: the light green object contains some target pieces, with P=0.39, R=0.15, and not removed by our FA detector. The orange object has intensities similar to that of clay targets and was not rejected by our FA detector.

Strengths and Weaknesses

- Strengths
  - Achieved good results (PD around 90% and PFA around 10%) with just two operations: pixel classification (segmentation) and false alarm detection (region classification)
  - Our techniques do not require metal artifact reduction, bulk/sheet classification, special processing, shape information, and split-and-merge processing
  - Our techniques (MRF-EM and collective FA classification) are not seen in prior patent literature
  - Relatively robust to moderate to medium metal artifacts
- Weaknesses
  - Targets can be still be merged with nearby/attaching objects with similar intensity distributions, causing them to be missed (high recall, low precision)

46

Future Improvement

- Objective: split some of the merged targets
- Potential solution 1: use edge information: neighboring objects with similar intensity distribution generally still have edges between them
- Caution: using edges may split targets that have edges from metal artifact
- Potential solution 2: detect regions that need to be split
- Caution: only a small number of regions (around 30) among all regions (several hundred) needs to be split
- Potential solution 3: make our algorithm more adaptive to the specifics of an image, e.g., amount of metal artifacts

47

Comments on Data

- The images are a diverse mix with some very interesting and challenging cases
- The ground truth set is good
- The scoring tools are very good
- We plan to use this beyond this project
  - Looking into more drastically different segmentation algorithms
  - Investigate how better reconstruction can improve segmentation and classification results (PD/PFA)

48

What you learned

- The images are very diverse and complex
- Solving a problem for one image may create problems for other images

49

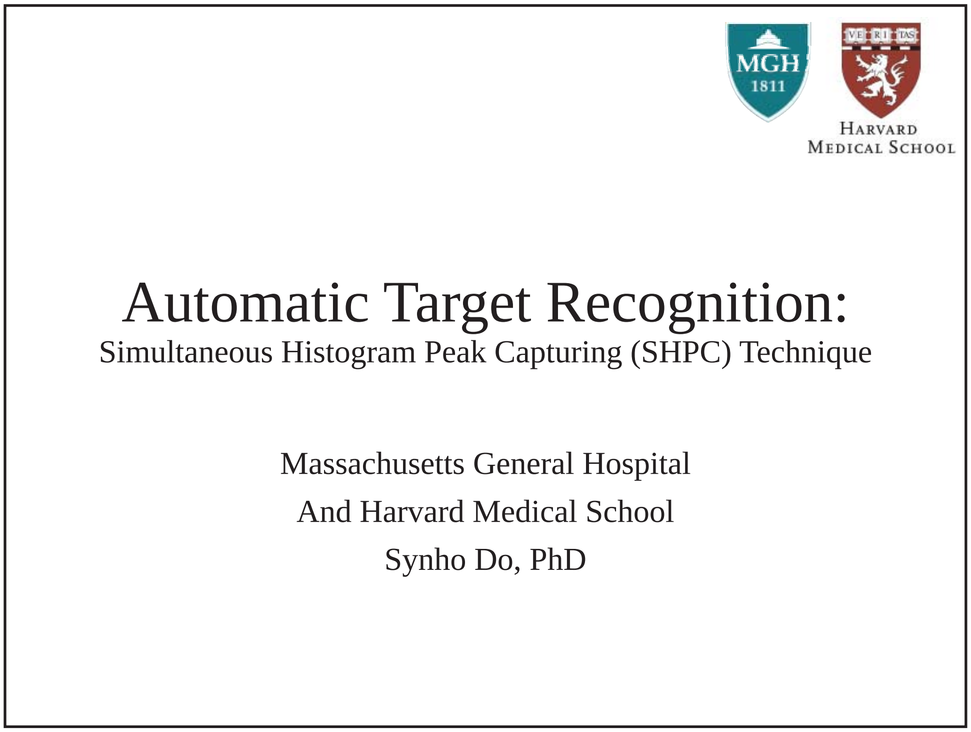
Another Result: 90%PD, 9%PFA

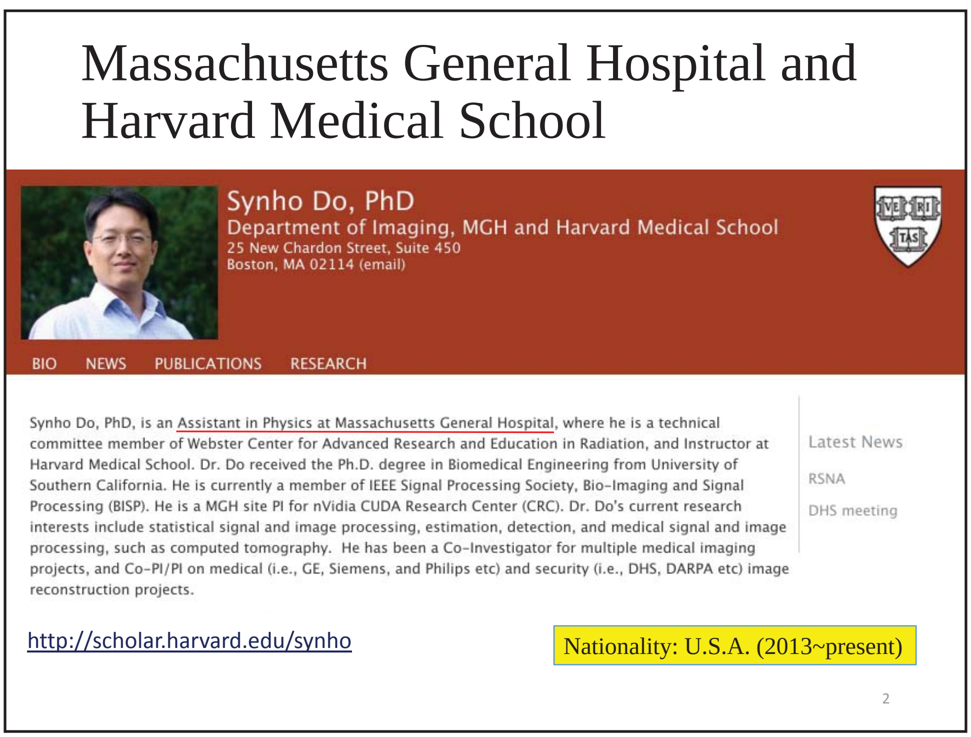
Target Type	Target Subtype or Form	Level of Difficulty	Num Targets	Num Detected	PD [%]
Target	All	All	407	366	89.9
Target	Clay	All	111	106	95.5
Target	Rubber	All	158	143	90.5
Target	Saline	All	138	117	84.8
Target	Bulk	All	270	237	87.8
Target	Sheet	All	137	129	94.2
Target	All	Low	77	75	97.4
Target	Clay	Low	29	29	100
Target	Rubber	Low	22	21	95.5
Target	Saline	Low	26	25	96.2
Target	Bulk	Low	56	55	98.2
Target	Sheet	Low	21	20	95.2
Target	All	High	317	280	88.3
Target	Clay	High	82	77	93.9
Target	Rubber	High	125	113	90.4
Target	Saline	High	110	90	81.8
Target	Bulk	High	201	171	85.1
Target	Sheet	High	116	109	94
Pseudo-target	Sheet	High	10	0	0
Num Non-targets	Num FAs	PFA [%]			
1370	128	9.3			

50

11.5.2.6 ATR Development – Do

"Automatic Target Recognition: Simultaneous Histogram Peak Capturing (SHPC) Technique"





PD/PFA results

(Ntarget=93, HUBW=8, AD=1500)

Target Type	Target Subtype or Form	Level of Difficulty	Num Targets	Num Detected	PD [%]
Target	All	All	407	384	94.3
Target	Clay	All	111	104	93.7
Target	Rubber	All	158	149	94.3
Target	Saline	All	138	131	94.9
Target	Bulk	All	270	253	93.7
Target	Sheet	All	137	131	95.6
Target	All	Low	77	74	96.1
Target	Clay	Low	29	29	100
Target	Rubber	Low	22	19	86.4
Target	Saline	Low	26	26	100
Target	Bulk	Low	56	55	98.2
Target	Sheet	Low	21	19	90.5
Target	All	High	317	299	94.3
Target	Clay	High	82	75	91.5
Target	Rubber	High	125	121	96.8
Target	Saline	High	110	103	93.6
Target	Bulk	High	201	187	93
Target	Sheet	High	116	112	96.6
Pseudo-target	Sheet	High	10	10	100
Num Non-targets	Num FAs	PFA [%]			
1371	114	8.3			

3

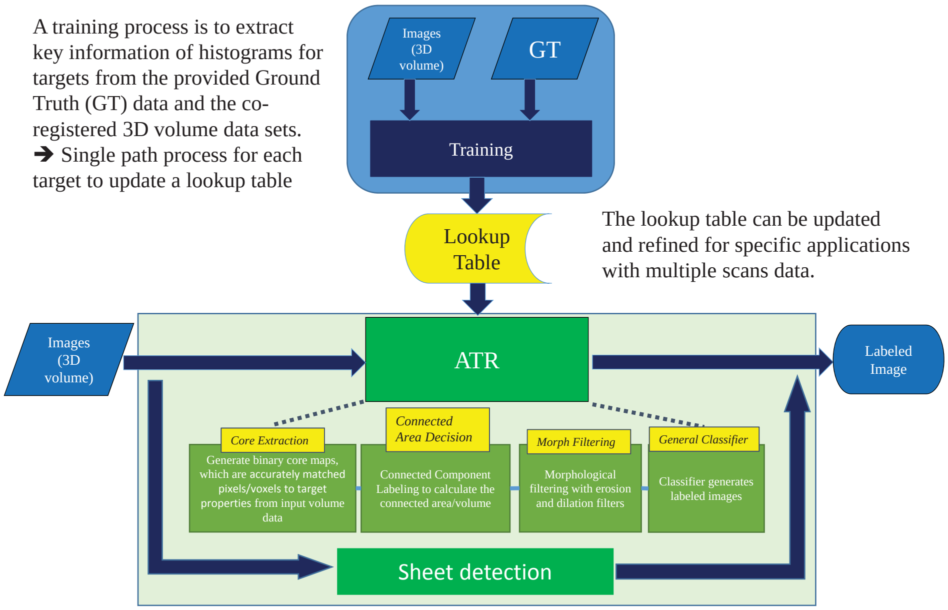
Simultaneous Histogram Peak Capturing (SHPC) Technique

Motivations:

- Minimum algorithm training is necessary for real applications.
- Shape independent feature selection is necessary.
- Quantitative measurement based algorithm development is necessary
  ➔ Target specific parameter extraction is necessary.
  - Material specific intrinsic HU value selection, HU Bandwidth selection, and target core volume decision
- Minimum volume consideration is necessary (from the target specifications.)
- Thin sheet detection algorithm to segment a defined thickness of objects.
- Simple method to add or subtract target materials is necessary (For independent training and tuning for each target.)
- Easy implementation is optional:
  - Easy to parallelize (i.e., GPU, Multi-core CPUs etc)
  - Easy to adjust for specific purposes (High density material detection, Organic material detection, Explosive detection, and Liquid detection, etc)
  - Individual target PD/PFA adjustment is possible.
- Potential to be extended to multi-spectral data ATR

4

Algorithm Flow Chart (1/2)



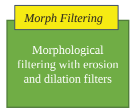
Core Extraction

- To generate binary core maps, which are accurately matched pixels/voxels to target material specific HU values.
- To generate a robust intermediate result even in the poor image qualities. (i.e., Blurred boundary, non-colored noise, Metal artifacts etc.)
- Utilize a Peak HU value and a HU Band width, which is a similar to the contrast window used in the medical field. In our case, the HU Bandwidth for each target is narrow and variable.

Connected Area Decision

- To improve the core extraction results, it calculates connected component area/volume.
- To remove non-core pixels/voxels, which are extracted from the previous core extraction process, it uses a threshold to keep the cores and to remove the non-cores .
- It generates multiple binary maps with the total number of connected components , pixel/voxel index lists for each connected component, and area/volume calculations.

6



Morphological Filtering

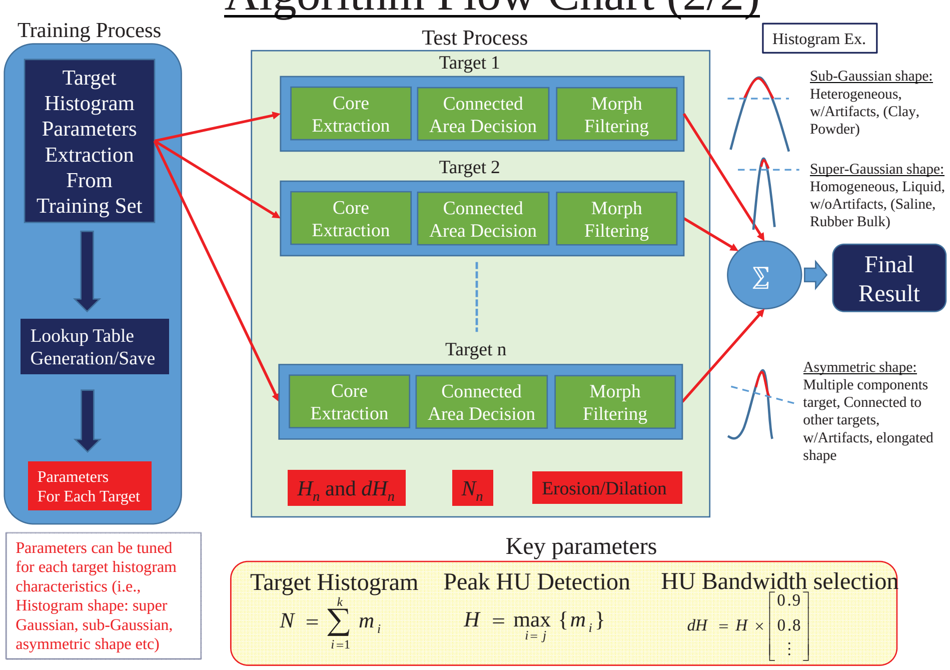
- To clean up boundaries of targets, it uses a morphological erosion filter and a dilation filter. Since the boundaries of targets don't keep the intrinsic HU value of the target and the narrow HU bandwidth allow the bumpy boundaries, it is very effective to make the target boundary smooth.
- In addition, the morphological dilation filter fills the small holes inside of a target object.
- For morphological filters, disks were tested in 2D simulations and 3D balls were implemented for 3D real cases.

General Classifier

- To generate labeled output, we considered to implement a non-statistical classifier. In this case, we don't need to build a statistical model, which have a generalization problem and requires a complicated training process, and use a look-up table method.
- A labeling matrix generated from the segmented structure was used to assign numbers to the objects based on those sizes.

7

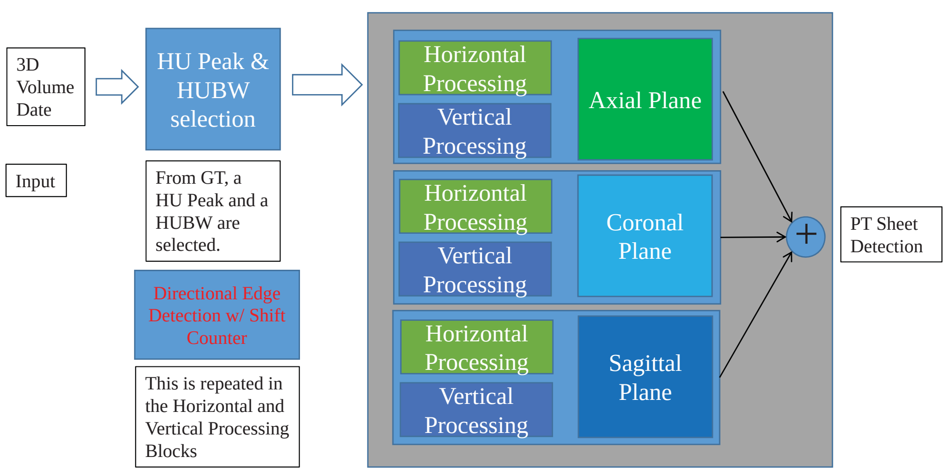
Algorithm Flow Chart (2/2)



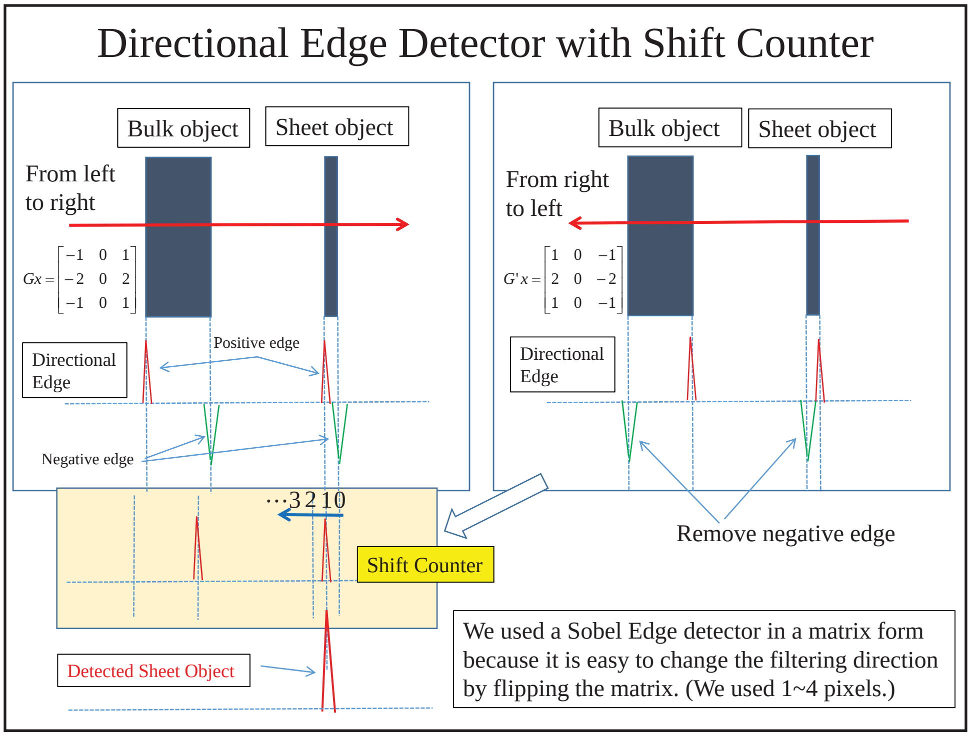
Glossary

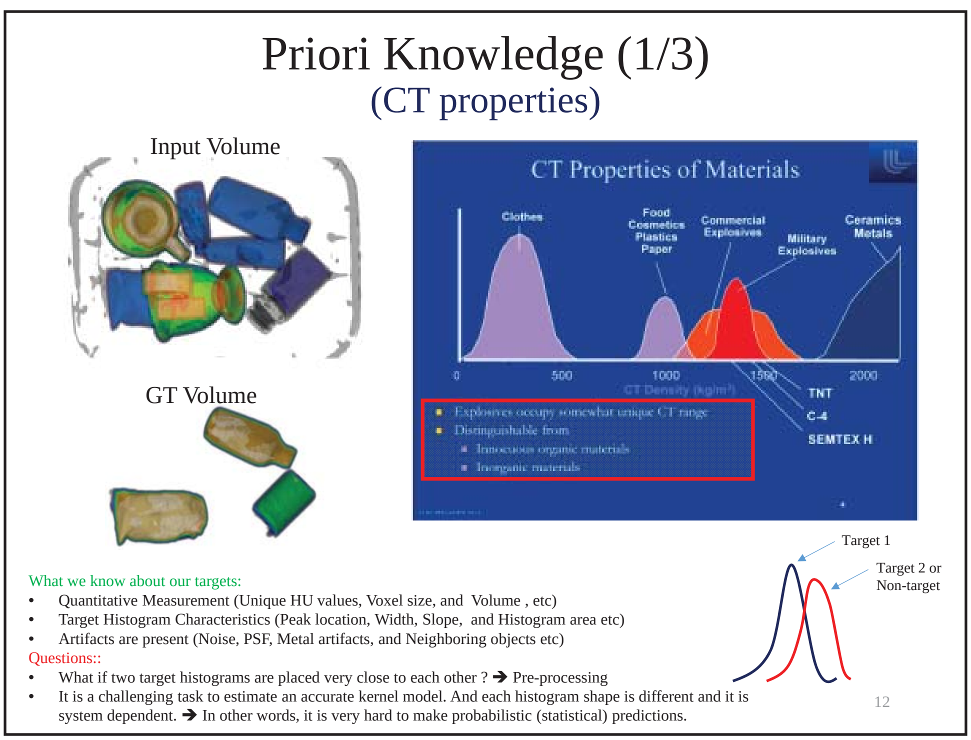
- Core extraction: It is the first step of the proposed method to find the core (accurately matched pixels or voxels to target properties) of target. It uses a material specific HU value, which is measured from a nominal target histogram, and a HU bandwidth (HUBW), which can be determined by users.
- Connected area detection: To make the core extraction more robust, we remove non-core voxels by using a connected area detection process. A pre-defined threshold value is used to keep the volumes only satisfying the minimum volume requirement.
- Morph filtering: The morphological erosion filter cleans up residuals of previous processing steps and the dilation filter can restore the size of target volume.
- Target histogram: A target histogram for a specific target can be measured by using multiple cases from the ground truth data set (like our case) or from repeated target scans. The measured target histogram will be saved as a look-up table so that we could use it for test cases. (i.e., $N = \sum_{i=1}^{k} m_i$)
- Peak detection: To extract the core of a target, we measure the HU value that gives the maximum bin count in the histogram. (i.e., $H = \max_{i=j} \{m_i\}$)
- Bandwidth selection: The selection of HUBW is related to how much deviation of HU we could allow to find the core of a target reliably. (i.e., $dH = H \times \begin{bmatrix} 0.9 \\ 0.8 \\ \vdots \end{bmatrix}$)

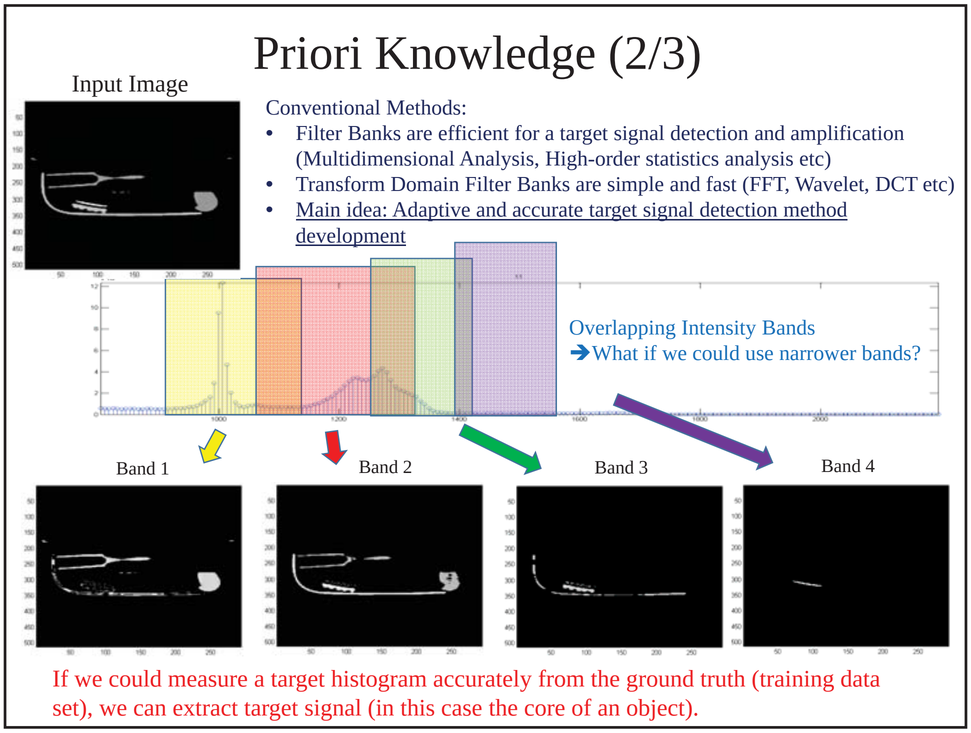
Sheet Detection Algorithm

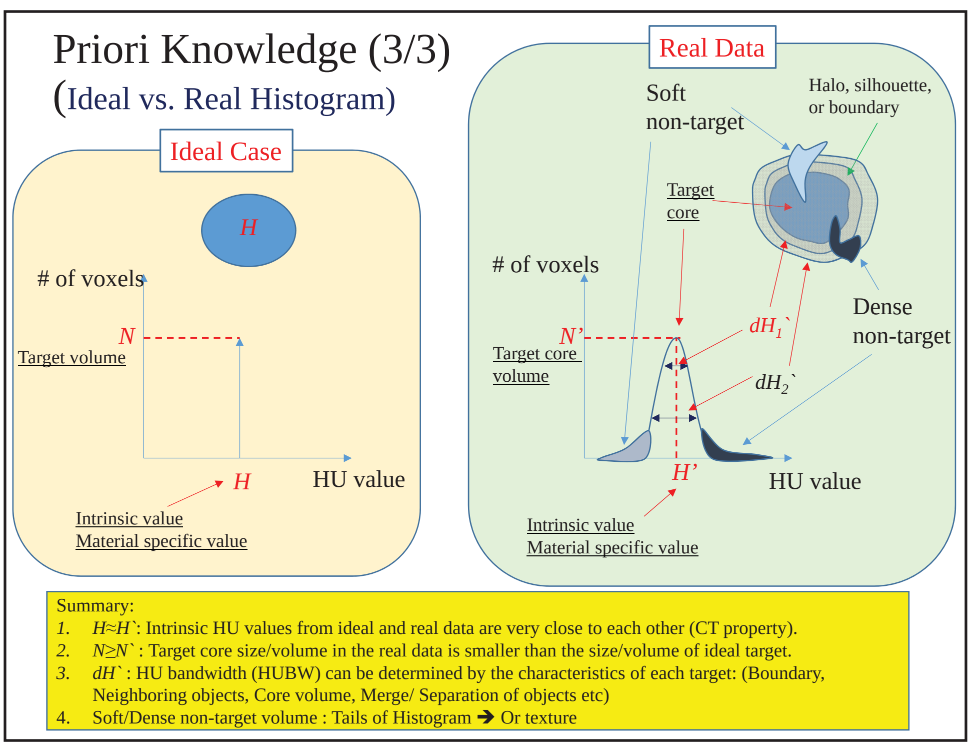


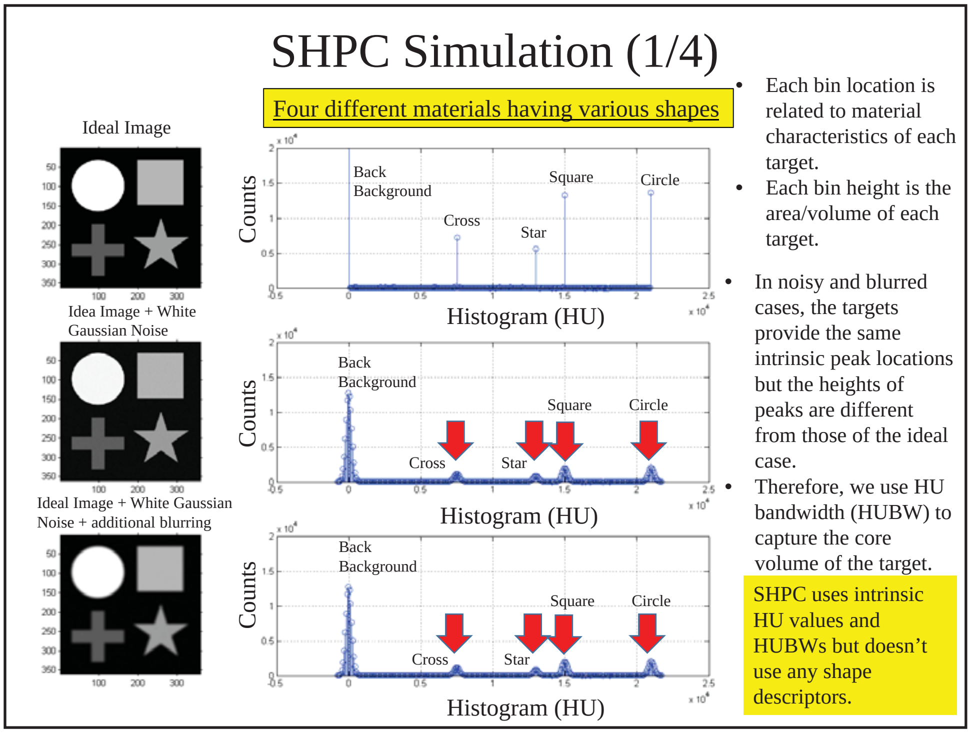
- PT sheets are thin and those occupy only a few pixels in 2D images.
- We use a directional edge detector with a shift counter to calculate the thickness of sheet objects.
- Horizontal and vertical processing blocks are the same process except edge detection directions.
- Axial, Coronal, and Sagittal planes are processed to combine for the final results.

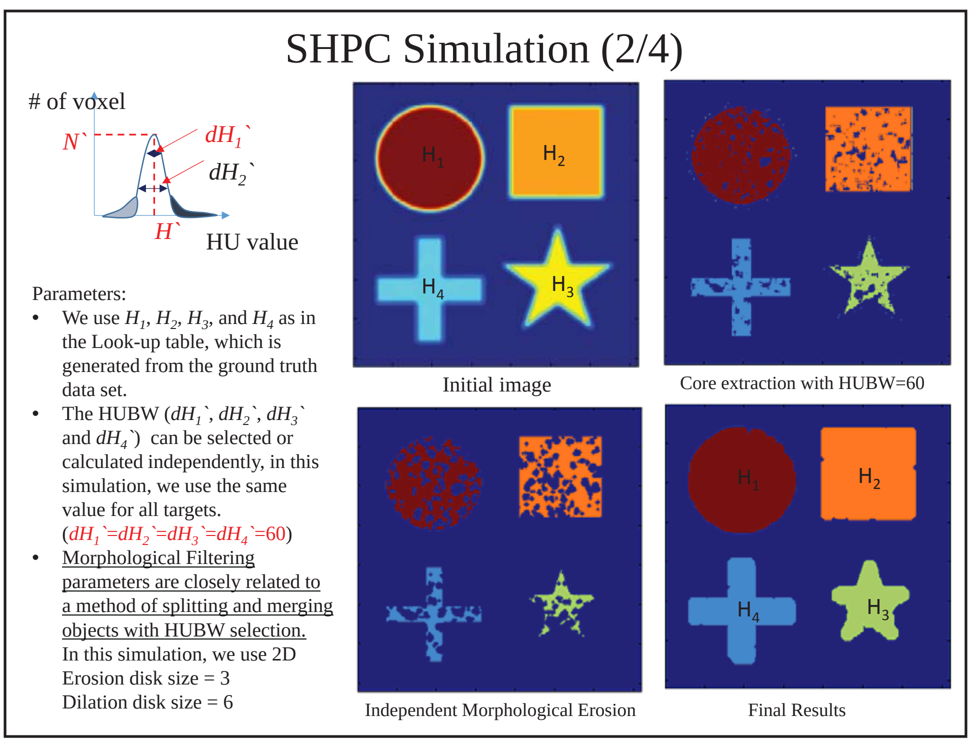

Directional Edge Detector with Shift Counter
Bulk object
Sheet object
From left to right
$Gx = \begin{bmatrix} -1 & 0 & 1 \\ -2 & 0 & 2 \\ -1 & 0 & 1 \end{bmatrix}$
Positive edge
Directional Edge
Negative edge
Bulk object
Sheet object
From right to left
$G'x = \begin{bmatrix} 1 & 0 & -1 \\ 2 & 0 & -2 \\ 1 & 0 & -1 \end{bmatrix}$
Directional Edge
Remove negative edge
…3 2 1 0
Shift Counter
Detected Sheet Object
We used a Sobel Edge detector in a matrix form because it is easy to change the filtering direction by flipping the matrix. (We used 1~4 pixels.)

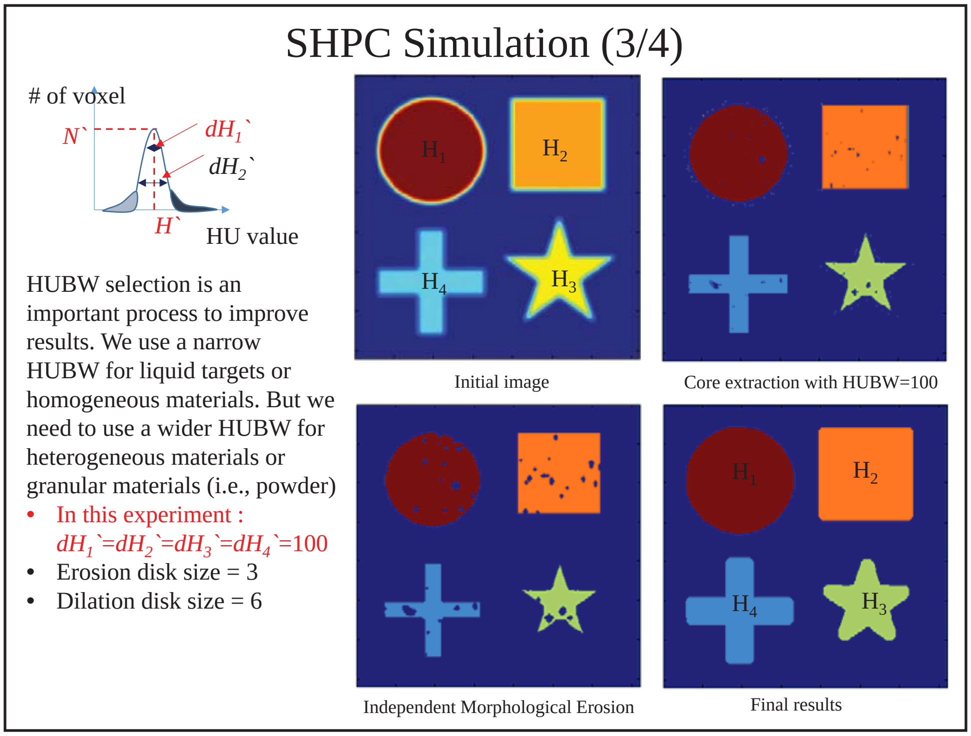

Priori Knowledge (1/3)
(CT properties)
Input Volume
GT Volume
CT Properties of Materials
Clothes
Food
Cosmetics
Plastics
Paper
Commercial
Explosives
Military
Explosives
Ceramics
Metals
0
500
1000
1500
2000
CT Density (kg/m³)
TNT
C-4
SEMTEX H
Explosives occupy somewhat unique CT range
Distinguishable from
Innocuous organic materials
Inorganic materials
Target 1
Target 2 or Non-target
What we know about our targets:
• Quantitative Measurement (Unique HU values, Voxel size, and Volume , etc)
• Target Histogram Characteristics (Peak location, Width, Slope, and Histogram area etc)
• Artifacts are present (Noise, PSF, Metal artifacts, and Neighboring objects etc)
Questions::
• What if two target histograms are placed very close to each other ? → Pre-processing
• It is a challenging task to estimate an accurate kernel model. And each histogram shape is different and it is system dependent. → In other words, it is very hard to make probabilistic (statistical) predictions.
12

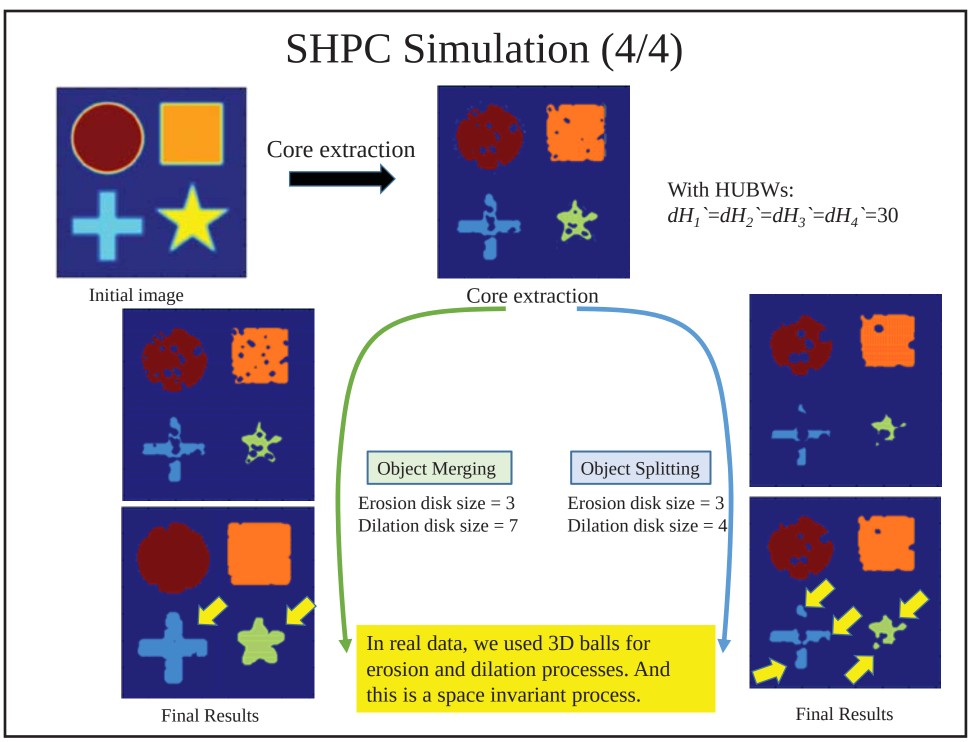

Priori Knowledge (2/3)
Input Image
Conventional Methods:
• Filter Banks are efficient for a target signal detection and amplification (Multidimensional Analysis, High-order statistics analysis etc)
• Transform Domain Filter Banks are simple and fast (FFT, Wavelet, DCT etc)
• Main idea: Adaptive and accurate target signal detection method development
Overlapping Intensity Bands
➔What if we could use narrower bands?
Band 1
Band 2
Band 3
Band 4
If we could measure a target histogram accurately from the ground truth (training data set), we can extract target signal (in this case the core of an object).


Priori Knowledge (3/3)
(Ideal vs. Real Histogram)
Ideal Case
H
# of voxels
N
Target volume
H
HU value
Intrinsic value
Material specific value
Real Data
Soft
non-target
Halo, silhouette,
or boundary
Target
core
# of voxels
N'
Target core
volume
dH1`
dH2`
Dense
non-target
H'
HU value
Intrinsic value
Material specific value
Summary:
1. H≈H`: Intrinsic HU values from ideal and real data are very close to each other (CT property).
2. N≥N` : Target core size/volume in the real data is smaller than the size/volume of ideal target.
3. dH` : HU bandwidth (HUBW) can be determined by the characteristics of each target: (Boundary, Neighboring objects, Core volume, Merge/ Separation of objects etc)
4. Soft/Dense non-target volume : Tails of Histogram ➔ Or texture


SHPC Simulation (1/4)
Four different materials having various shapes
Ideal Image
Idea Image + White Gaussian Noise
Ideal Image + White Gaussian Noise + additional blurring
Counts
Histogram (HU)
Back Background
Cross
Star
Square
Circle
Each bin location is related to material characteristics of each target.
Each bin height is the area/volume of each target.
In noisy and blurred cases, the targets provide the same intrinsic peak locations but the heights of peaks are different from those of the ideal case.
Therefore, we use HU bandwidth (HUBW) to capture the core volume of the target.
SHPC uses intrinsic HU values and HUBWs but doesn't use any shape descriptors.


SHPC Simulation (2/4)
# of voxel
N`
dH1`
dH2`
H`
HU value
Parameters:
We use H1, H2, H3, and H4 as in the Look-up table, which is generated from the ground truth data set.
The HUBW (dH1`, dH2`, dH3` and dH4`) can be selected or calculated independently, in this simulation, we use the same value for all targets.
(dH1`=dH2`=dH3`=dH4`=60)
Morphological Filtering parameters are closely related to a method of splitting and merging objects with HUBW selection.
In this simulation, we use 2D
Erosion disk size = 3
Dilation disk size = 6
H1
H2
H3
H4
Initial image
Core extraction with HUBW=60
Independent Morphological Erosion
Final Results

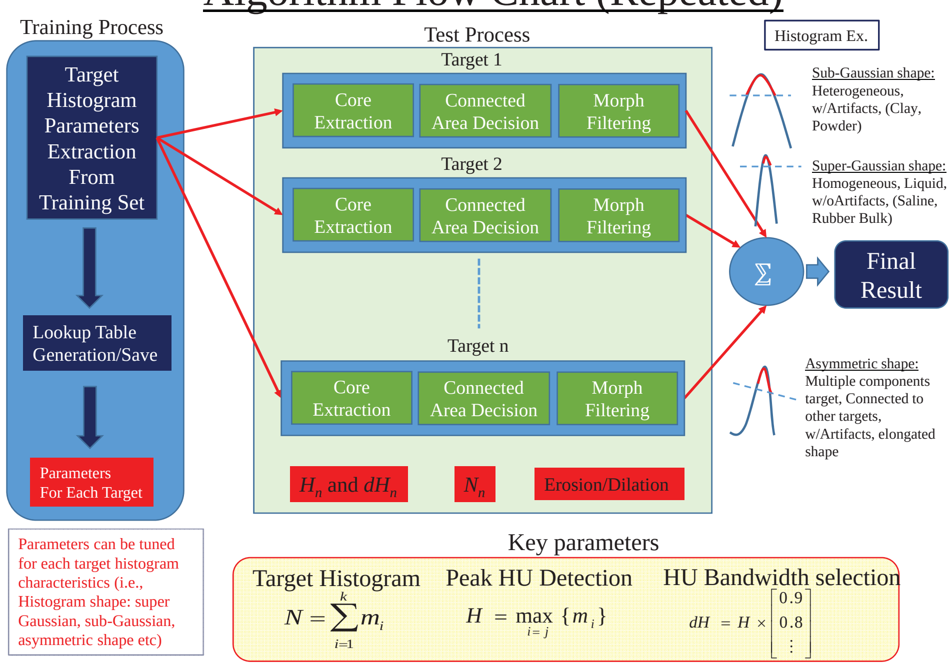

SHPC Simulation (3/4)
# of voxel
$N`$
$dH_1`$
$dH_2`$
$H`$
HU value
HUBW selection is an important process to improve results. We use a narrow HUBW for liquid targets or homogeneous materials. But we need to use a wider HUBW for heterogeneous materials or granular materials (i.e., powder)
• In this experiment : $dH_1`=dH_2`=dH_3`=dH_4`=100$
• Erosion disk size = 3
• Dilation disk size = 6
H_1
H_2
H_4
H_3
Initial image
Core extraction with HUBW=100
Independent Morphological Erosion
Final results

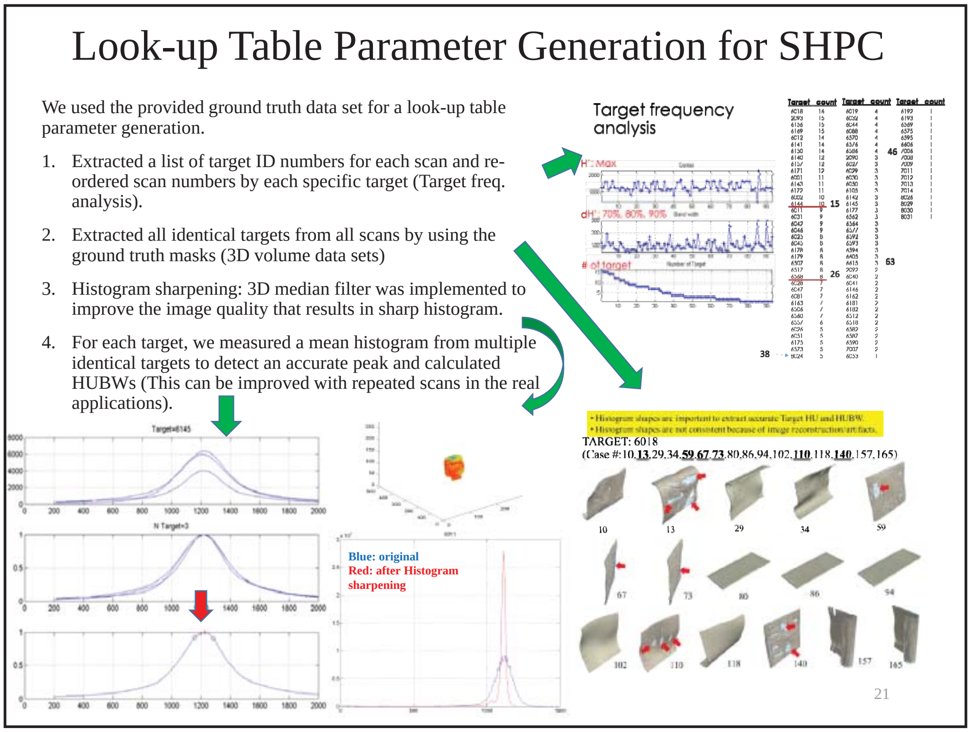

SHPC Simulation (4/4)
Core extraction
Initial image
Core extraction
With HUBWs:
$dH_1`=dH_2`=dH_3`=dH_4`=30$
Object Merging
Erosion disk size = 3
Dilation disk size = 7
Object Splitting
Erosion disk size = 3
Dilation disk size = 4
In real data, we used 3D balls for erosion and dilation processes. And this is a space invariant process.
Final Results
Final Results

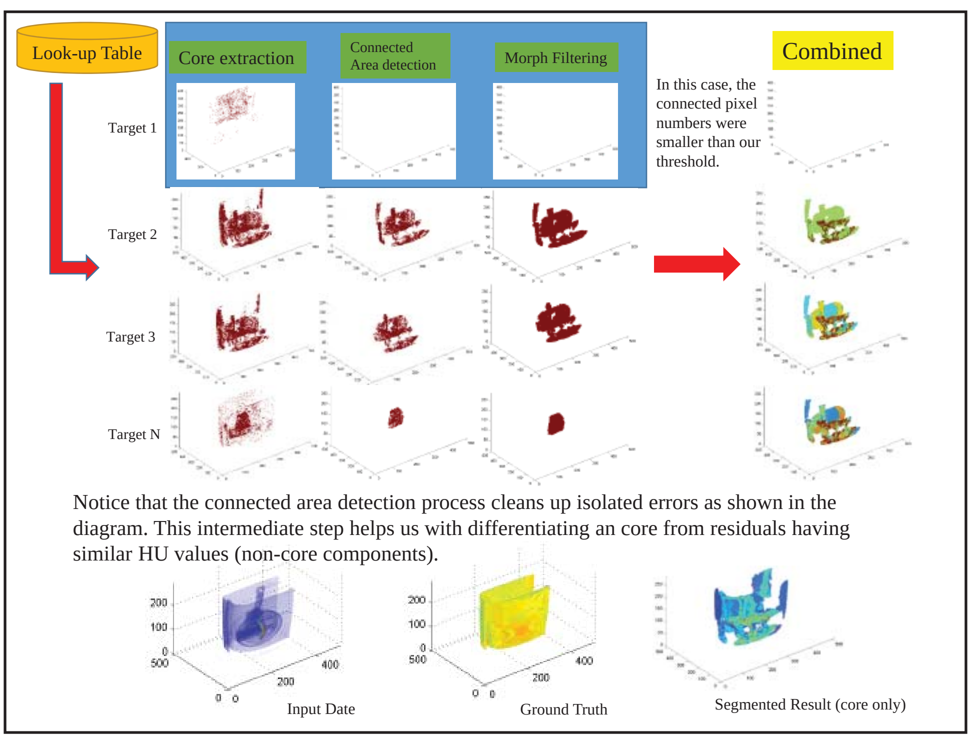
Real Data Cases

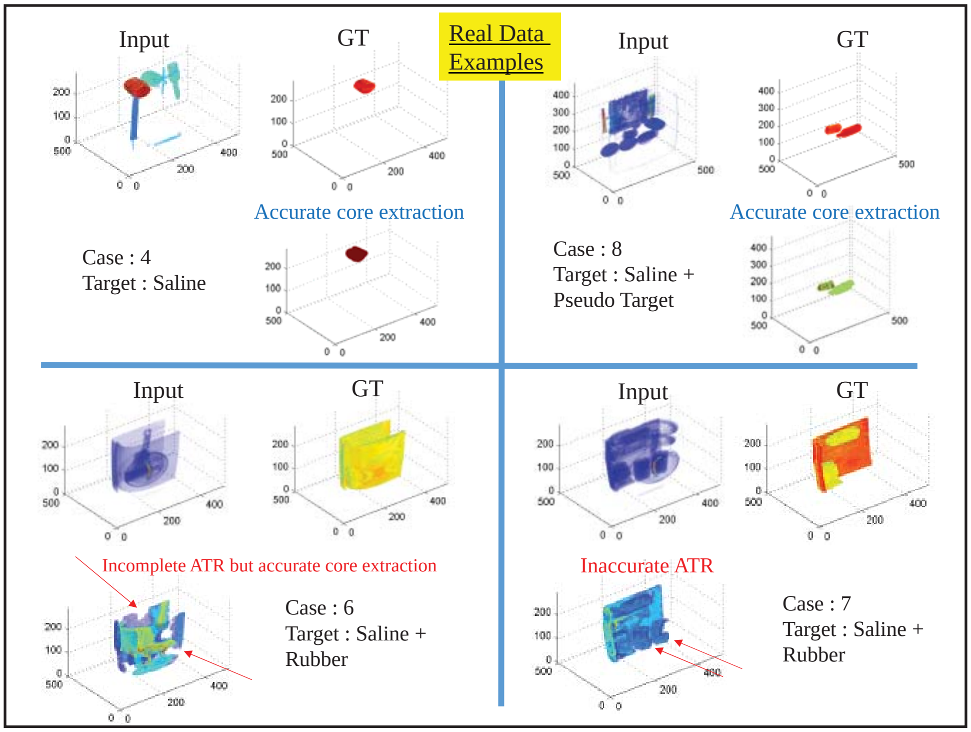
19

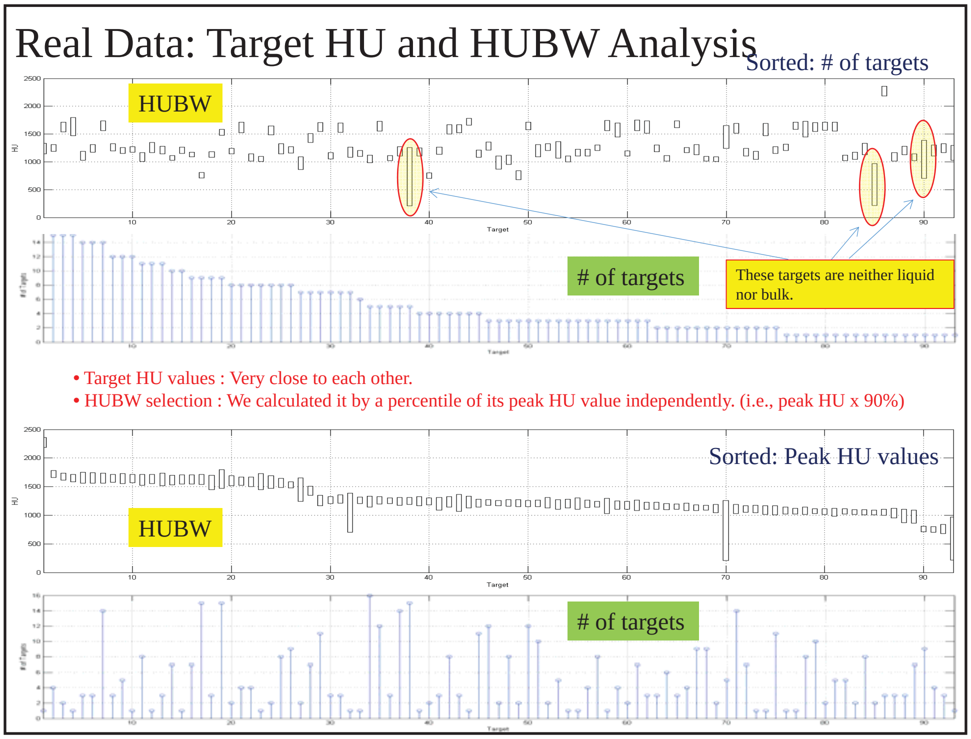
Algorithm Flow Chart (Repeated)

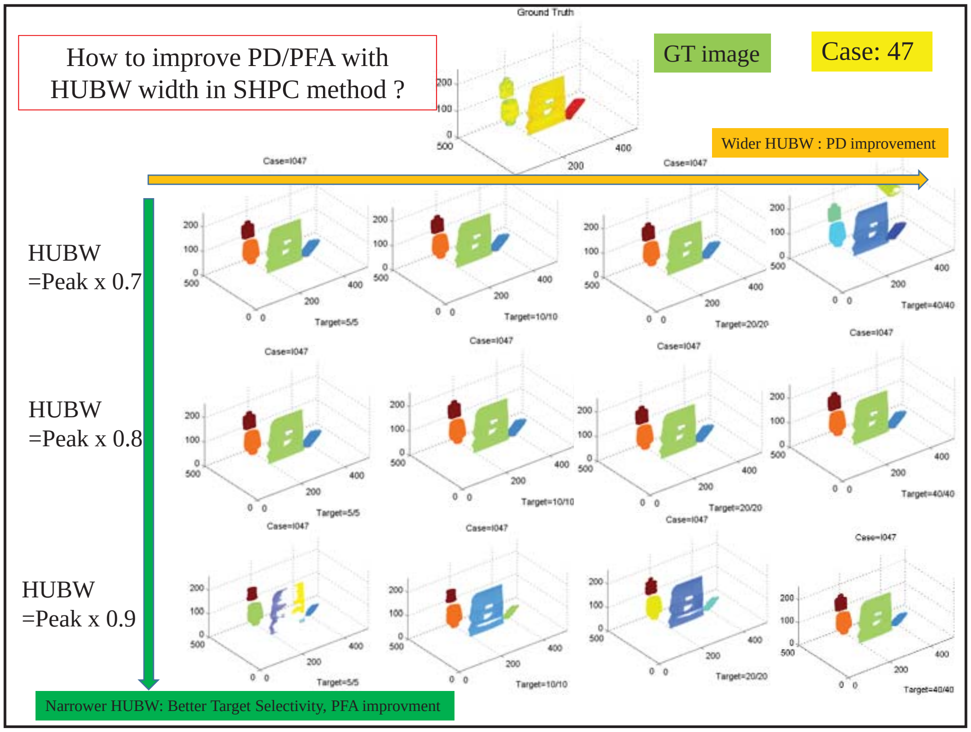


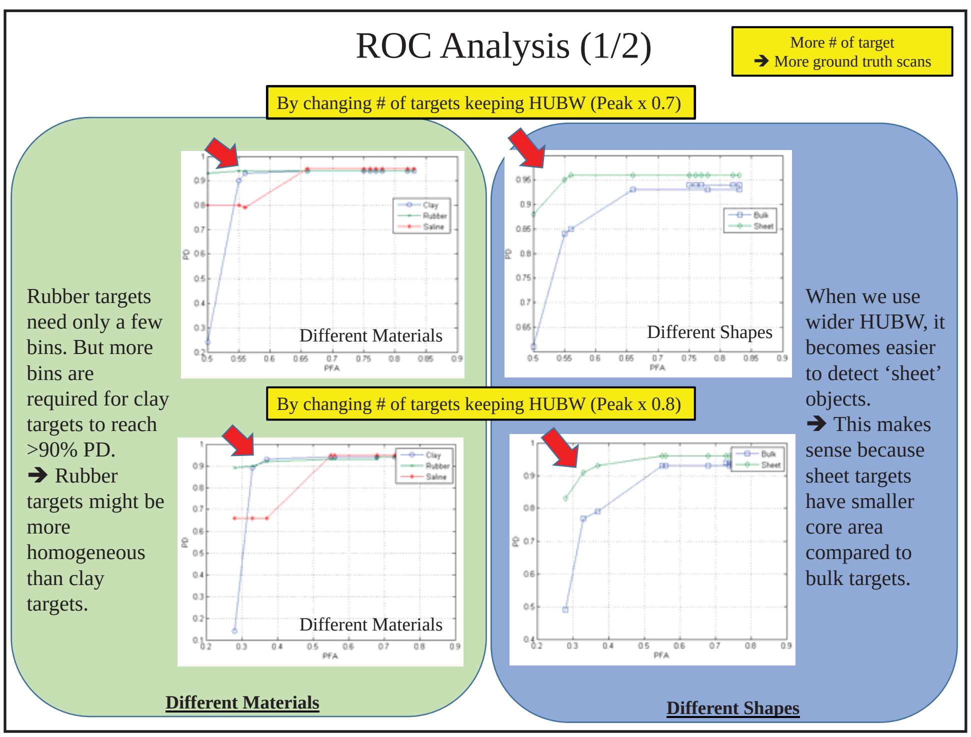

Look-up Table Parameter Generation for SHPC
We used the provided ground truth data set for a look-up table parameter generation.
1. Extracted a list of target ID numbers for each scan and re-ordered scan numbers by each specific target (Target freq. analysis).
2. Extracted all identical targets from all scans by using the ground truth masks (3D volume data sets)
3. Histogram sharpening: 3D median filter was implemented to improve the image quality that results in sharp histogram.
4. For each target, we measured a mean histogram from multiple identical targets to detect an accurate peak and calculated HUBWs (This can be improved with repeated scans in the real applications).
Target frequency analysis
H: Max
dH: 70%, 80%, 90%
# of target
• Histogram shapes are important to extract accurate Target HU and HUBW.
• Histogram shapes are not consistent because of image reconstruction artifacts.
TARGET: 6018
(Case #:10,13,29,34,59,67,73,80,86,94,102,110,118,140,157,165)
Blue: original
Red: after Histogram sharpening
21

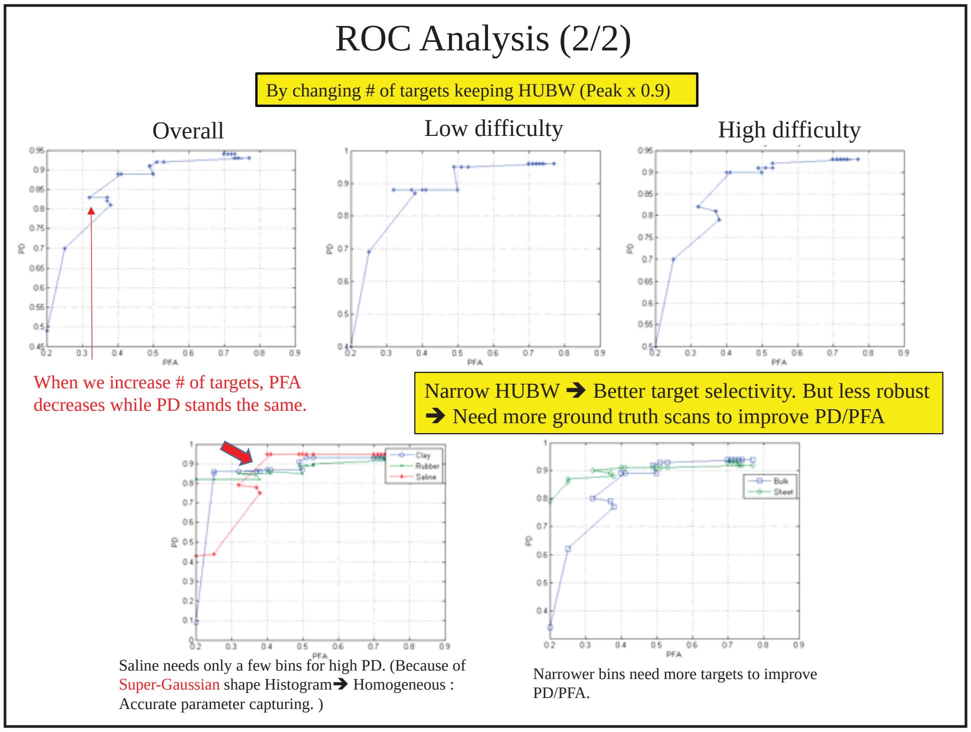

Look-up Table
Core extraction
Connected Area detection
Morph Filtering
Combined
Target 1
Target 2
Target 3
Target N
In this case, the connected pixel numbers were smaller than our threshold.
Notice that the connected area detection process cleans up isolated errors as shown in the diagram. This intermediate step helps us with differentiating an core from residuals having similar HU values (non-core components).
Input Date
Ground Truth
Segmented Result (core only)

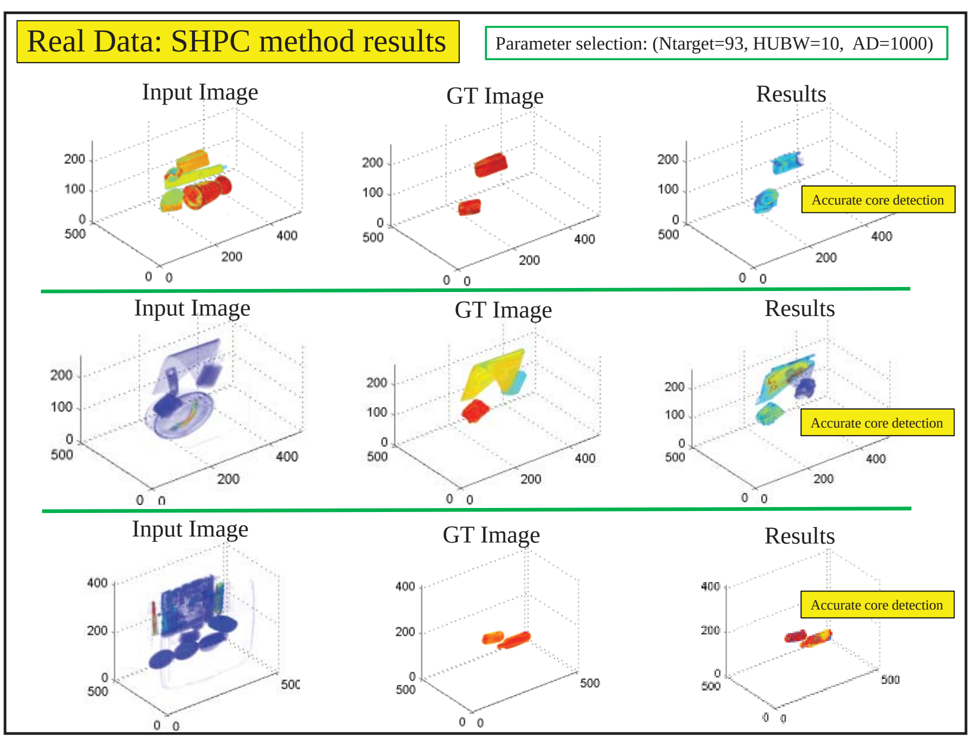

Real Data Examples
Input
GT
Accurate core extraction
Case : 4
Target : Saline
Input
GT
Accurate core extraction
Case : 8
Target : Saline + Pseudo Target
Input
GT
Incomplete ATR but accurate core extraction
Case : 6
Target : Saline + Rubber
Input
GT
Inaccurate ATR
Case : 7
Target : Saline + Rubber


Real Data: Target HU and HUBW Analysis
Sorted: # of targets
HUBW
# of targets
These targets are neither liquid nor bulk.
• Target HU values : Very close to each other.
• HUBW selection : We calculated it by a percentile of its peak HU value independently. (i.e., peak HU x 90%)
Sorted: Peak HU values
HUBW
# of targets

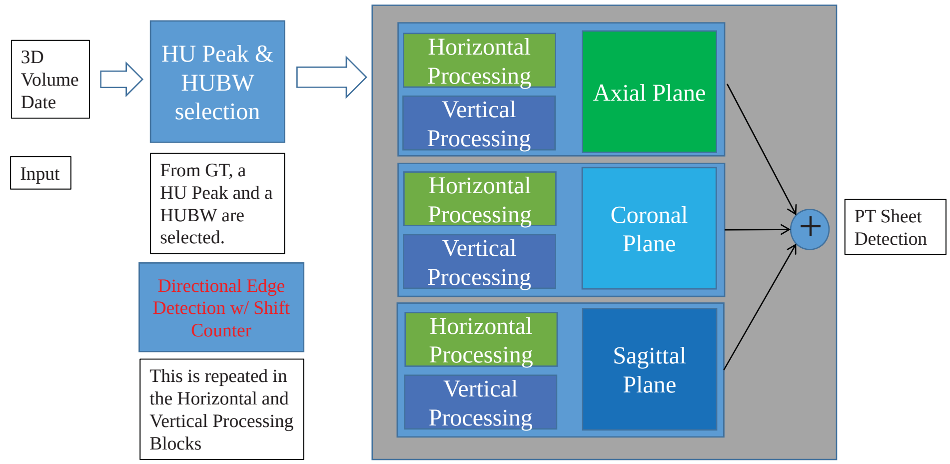

How to improve PD/PFA with HUBW width in SHPC method ?
GT image
Case: 47
Ground Truth
Wider HUBW : PD improvement
HUBW =Peak x 0.7
HUBW =Peak x 0.8
HUBW =Peak x 0.9
Target=5/5
Target=10/10
Target=20/20
Target=40/40
Case=I047
Narrower HUBW: Better Target Selectivity, PFA improvment


ROC Analysis (1/2)
More # of target
➔ More ground truth scans
By changing # of targets keeping HUBW (Peak x 0.7)
By changing # of targets keeping HUBW (Peak x 0.8)
Different Materials
Different Shapes
Clay
Rubber
Saline
Bulk
Sheet
PD
PFA
Rubber targets need only a few bins. But more bins are required for clay targets to reach >90% PD.
➔ Rubber targets might be more homogeneous than clay targets.
When we use wider HUBW, it becomes easier to detect 'sheet' objects.
➔ This makes sense because sheet targets have smaller core area compared to bulk targets.
Different Materials
Different Shapes


ROC Analysis (2/2)
By changing # of targets keeping HUBW (Peak x 0.9)
Overall
Low difficulty
High difficulty
PD
PFA
When we increase # of targets, PFA decreases while PD stands the same.
Narrow HUBW ➔ Better target selectivity. But less robust
➔ Need more ground truth scans to improve PD/PFA
Clay
Rubber
Saline
Bulk
Sheet
Saline needs only a few bins for high PD. (Because of Super-Gaussian shape Histogram➔ Homogeneous : Accurate parameter capturing.)
Narrower bins need more targets to improve PD/PFA.

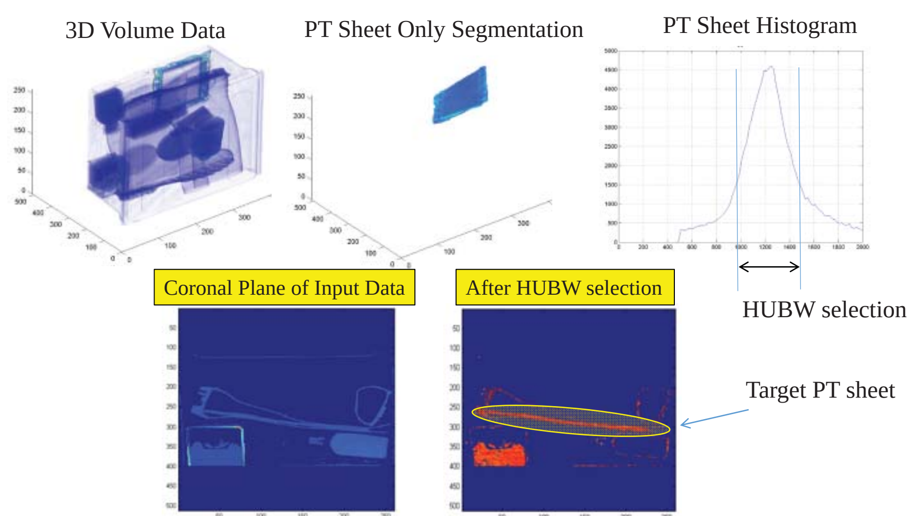

Real Data: SHPC method results
Parameter selection: (Ntarget=93, HUBW=10, AD=1000)
Input Image
GT Image
Results
Accurate core detection
Input Image
GT Image
Results
Accurate core detection
Input Image
GT Image
Results
Accurate core detection

Sheet Detection Algorithm

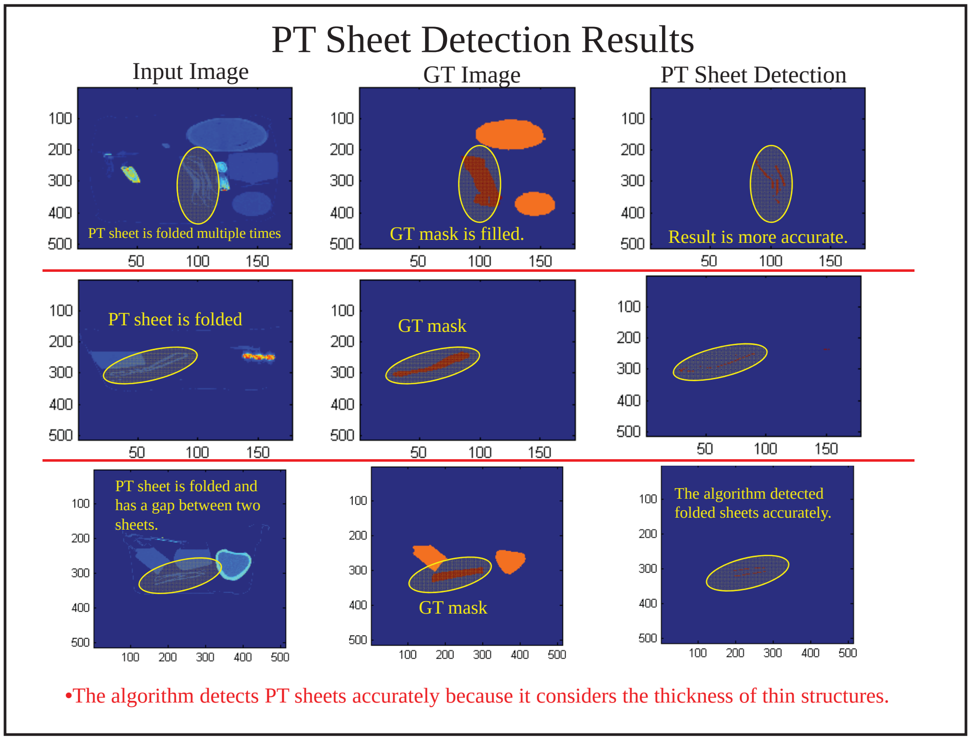


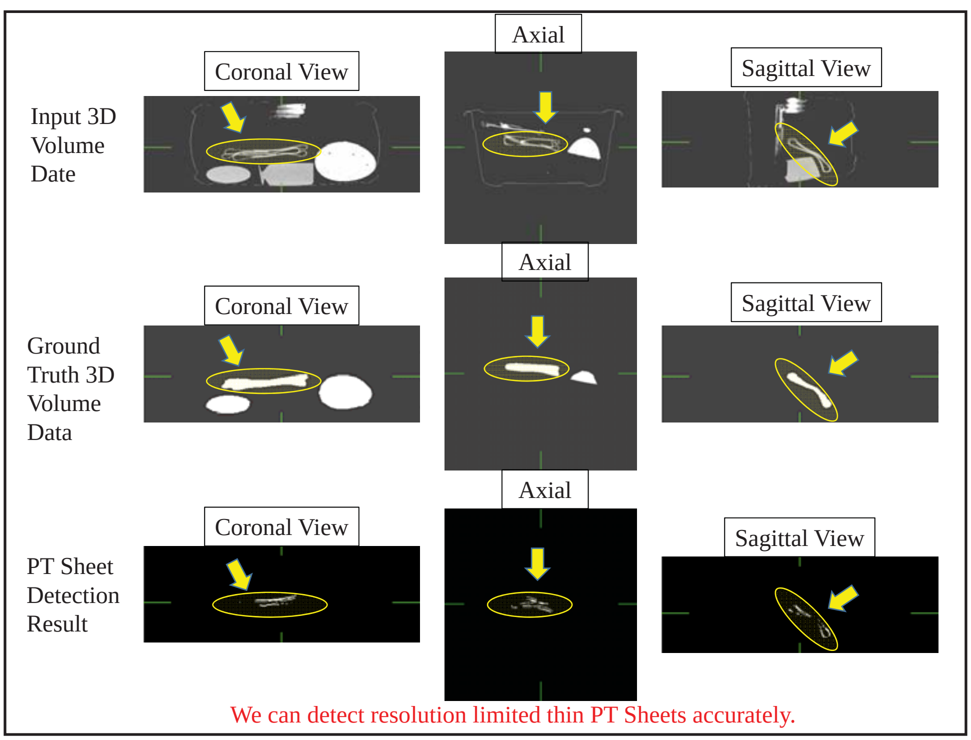
- PT sheets are thin and those occupy only a few pixels in 2D images.
- We use a directional edge detector with a shift counter to count the thickness of a sheet object.
- Horizontal and vertical processing blocks are the same process except edge detection directions.
- Axial, Coronal, and Sagittal planes are processed to combine for the final results.

PT Sheet Detection Process

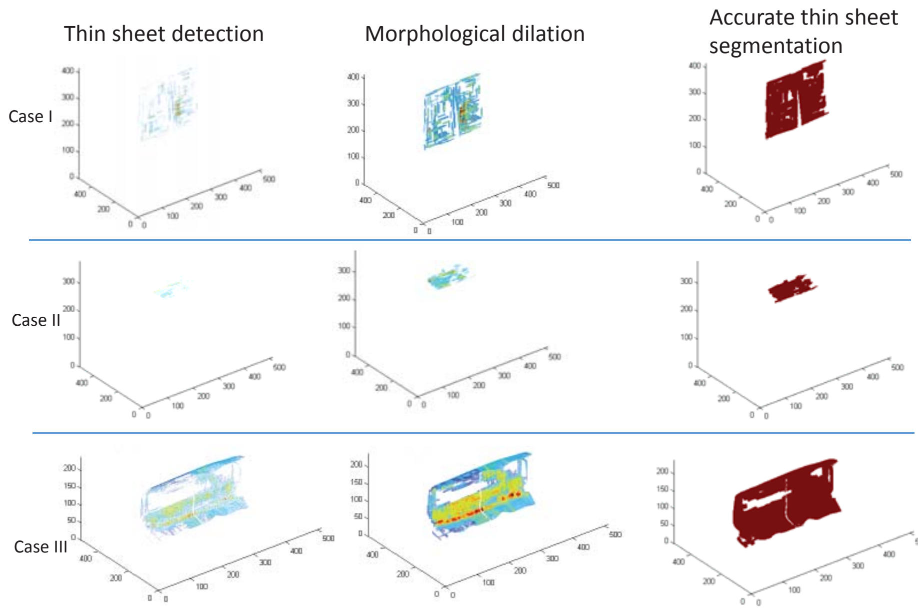


We are only interested in "sheet shape" in the selected HUBW. (We used 1000<x<1400 HU.)


PT Sheet Detection Results
Input Image
GT Image
PT Sheet Detection
PT sheet is folded multiple times
GT mask is filled.
Result is more accurate.
PT sheet is folded
GT mask
PT sheet is folded and has a gap between two sheets.
GT mask
The algorithm detected folded sheets accurately.
•The algorithm detects PT sheets accurately because it considers the thickness of thin structures.


Axial
Coronal View
Sagittal View
Input 3D Volume Date
Ground Truth 3D Volume Data
PT Sheet Detection Result
We can detect resolution limited thin PT Sheets accurately.

PT Sheet Detection Results



After thin sheet detection, a morphological filter was applied to increase the volume.

Sheet Detection Algorithm Summary

- The PT sheet detection algorithm takes into account the material and the thickness of sheet structures.
  - The target thickness can be controlled by a shift counter setting.
- It detects only sheet objects accurately.
- This algorithm is useful to detect resolution limited sheet structures.
- A morphological filtering can improve the connectivity of thin structures to visualize entire sheet objects.

PD/PFA results

(Ntarget=93, HUBW=8, AD=1500)

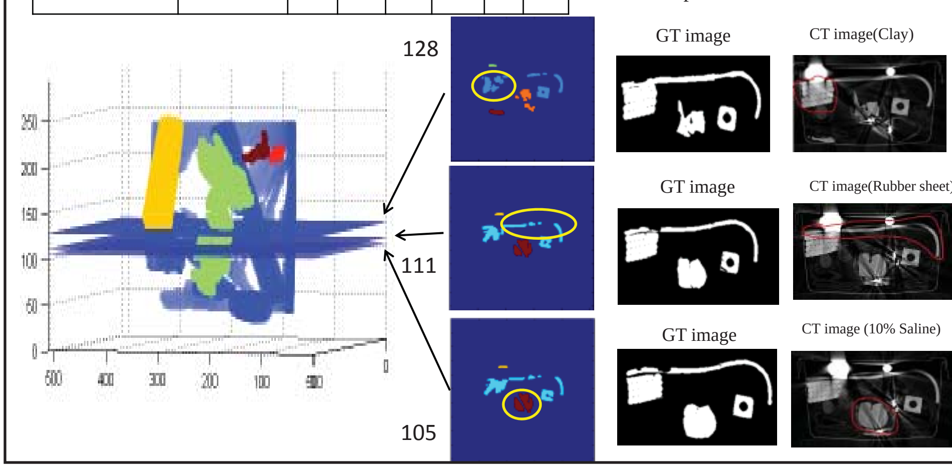
Target Type	Target Subtype or Form	Level of Difficulty	Num Targets	Num Detected	PD [%]
Target	All	All	407	384	94.3
Target	Clay	All	111	104	93.7
Target	Rubber	All	158	149	94.3
Target	Saline	All	138	131	94.9
Target	Bulk	All	270	253	93.7
Target	Sheet	All	137	131	95.6
Target	All	Low	77	74	96.1
Target	Clay	Low	29	29	100
Target	Rubber	Low	22	19	86.4
Target	Saline	Low	26	26	100
Target	Bulk	Low	56	55	98.2
Target	Sheet	Low	21	19	90.5
Target	All	High	317	299	94.3
Target	Clay	High	82	75	91.5
Target	Rubber	High	125	121	96.8
Target	Saline	High	110	103	93.6
Target	Bulk	High	201	187	93
Target	Sheet	High	116	112	96.6
Pseudo-target	Sheet	High	10	10	100
Num Non-targets	Num FAs	PFA [%]			
1371	114	8.3			

35

Presentation Cases (1,2, and 6)

Case	Target	ID	Mass	SSN	Loc	Ort	Slice
Bulk with bad streaks cased b y metal	Breast milk bag 10% Saline	6012	285	13	Bbb	z	105
Sheet with bad streaks cased b y metal, beam hardening and s catter	Rubber sheet 6.6mm	6018	685	13	Bcb	z	111
Bulk with bad shading caused by beam hardening and scatter	Clay	6051	286	13	Abb	z	128

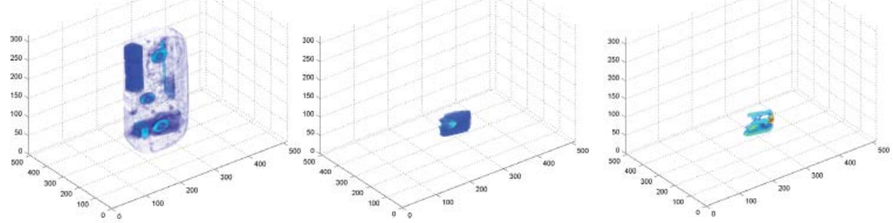
Clay was merged with a rubber sheet but 10% saline is segmented accurately even in the situation with streak artifacts.➔ This can be improved more with a warping based Metal Artifact Forecasting Method, which is shown in the backup slides.



Presentation Case (3/10)

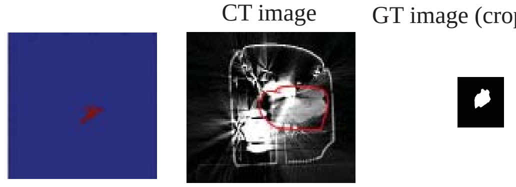
Case	Target	ID	Mass	SSN	Loc	Ort	Slice
Bulk inside elect ronics	Clay	6150	290	35	Aab	z	49

Input Image GT Image Results



CT image GT image (cropped)

Slice # = 49
Clay (Heterogeneous)

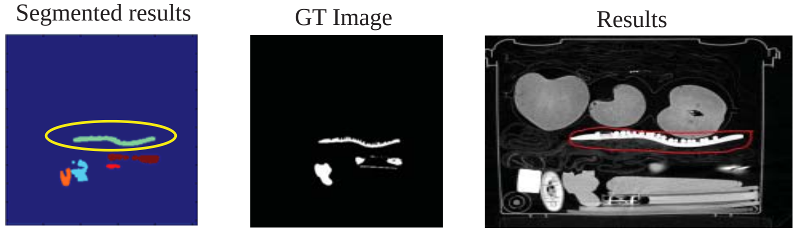


The core of clay target is detected accurately but the total Mass (290 cc) was not segmented perfectly. ➔ The Morphological dilation process can improve the results.

Presentation Case (4/10)

Case	Target	ID	Mass	SSN	Loc	Ort	Slice
Bulk with texture	Clay w/ glass beads	6193	410	193	bbc	x	198

Segmented results GT Image Results



The bulk with texture target (i.e., clay with glass beads) is segmented accurately even though the material is not homogenous, and has an elongated shape.
➔ The proposed algorithm could easily extract the core component, because the Mass is big as in this case (Mass =410 cc) and the image quality is good (no artifacts).

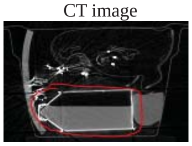
38

Presentation Case (5/10)

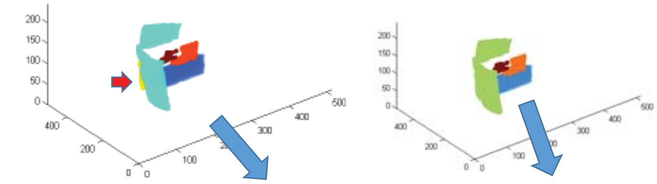
Case	Target	ID	Mass	SSN	Loc	Ort	Slice
Bulk with density close to water (~5% saline)	5% Saline – tin bottle	6163	274	63	baa	x	45

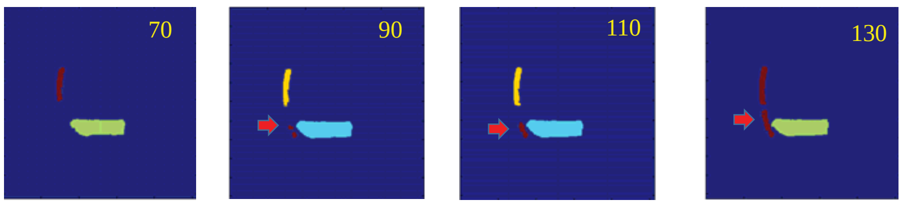
GT image

CT image



This case shows how target specific tuning can be executed for a rubber sheet without changing any parameters for the saline target.





Target specific tuning: HUBW = 70, 90, 110, and 130 (Rubber sheet), Saline (no change)

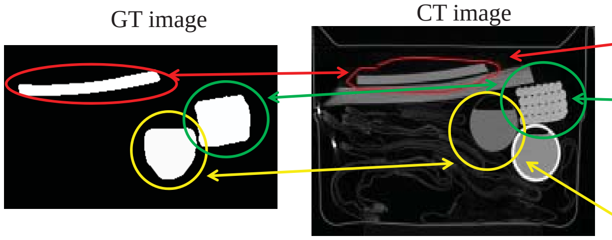
39

Presentation Case (7/10)

Case	Target	ID	Mass	SSN	Loc	Ort	Slice
Sheet laying on top of anot her flat object	3/8 rubber sheet on Ell e magazine	6144	345	33	Bca	x	46

GT image

CT image

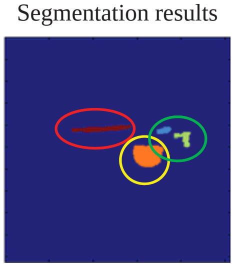


Rubber sheet (3/8 in.)

clay

5% Saline in the breast milk bottle

Segmentation results



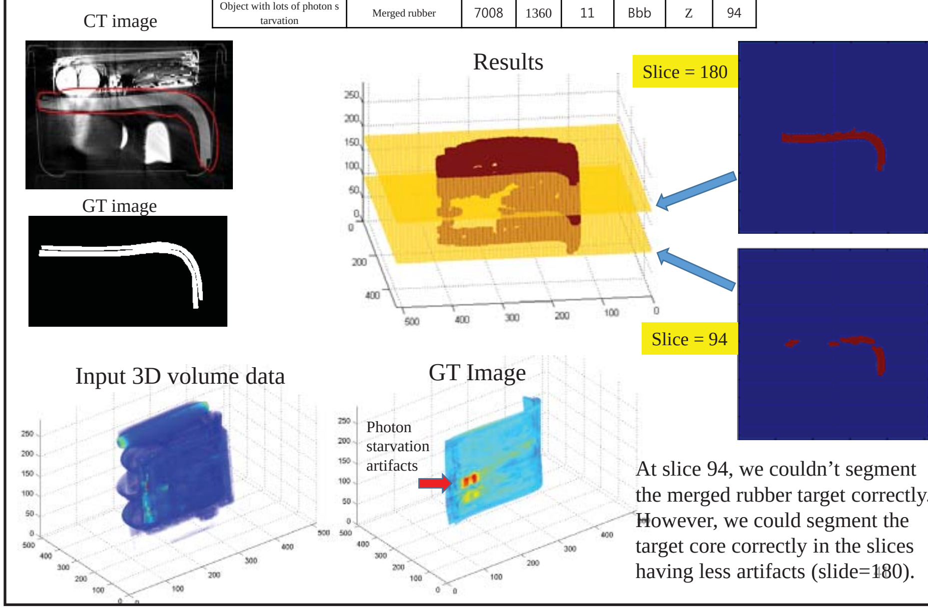
This presentation case shows us how the ground truth mask is important for the target detection in the test set.

- The core of Rubber sheet (3/8 in) was segmented accurately because the material is homogeneous and image has minimum artifacts.
- The liquid (5% saline) was detected accurately without any problem.
- The clay was not accurately segmented and was not merged, because the ground truth mask includes air pockets between the clay dough. It made the parameter selection process inaccurate.

40

Presentation Case (8/10)

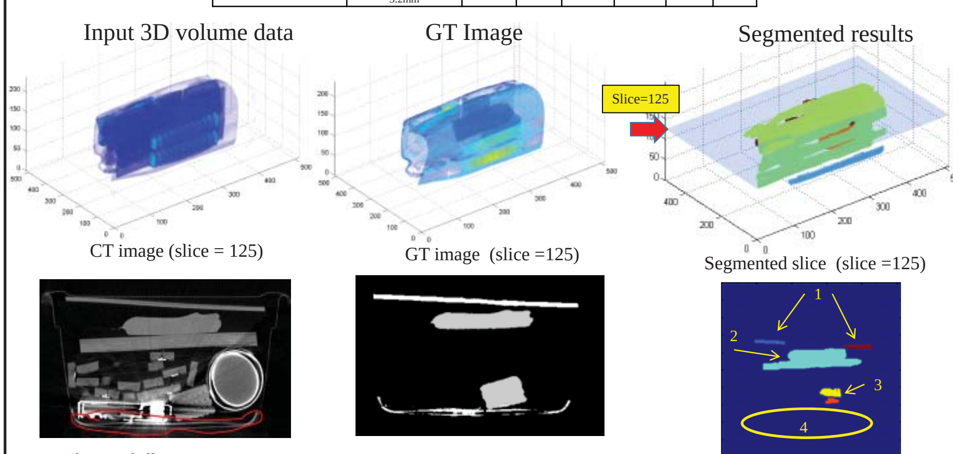
Case	Target	ID	Mass	SSN	Loc	Ort	Slice
Object with lots of photon s tarvation	Merged rubber	7008	1360	11	Bbb	Z	94



At slice 94, we couldn't segment the merged rubber target correctly. However, we could segment the target core correctly in the slices having less artifacts (slide=180).

Presentation Case (9/10)

Case	Target	ID	Mass	SSN	Loc	Ort	Slice
PT sheet based on thickness	Neoprene rubber sheet 3.2mm	8026	350	18	Bab	Z	125



This is a challenging case:

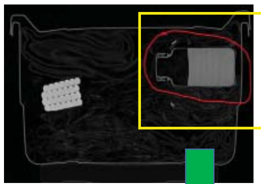
1. The target rubber sheet is splitted to two different objects.
2. A target (ceramic mud + clay) is merged with a neighboring object.
3. Clay is splitted because of streaking artifacts and segmented to two different objects. ➔ This can be resolved by a warping based Metal Artifact Forecasting method, attached in the backup slide.
4. A Pseudo-target (Neoprene rubber sheet 3.2 mm) is not detected because the artifacts are present and the pseudo-target is too thin to find the core.. ➔ Improved system resolution could overcome this problem.

42

Presentation Case (10/10)

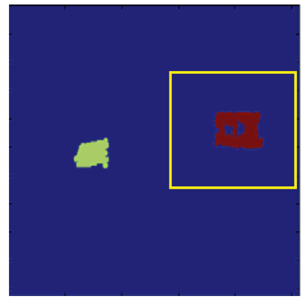
Case	Target	ID	Mass	SSN	Loc	Ort	Slice
PT Powder (based on densit y, not mass)	TA_MH01 plastic bottl e+ powder	6026	277	12	Cca	X	105

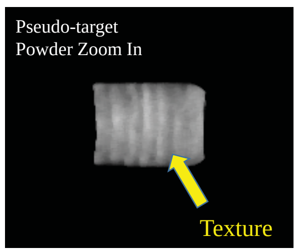
CT image



GT image

Segmented results

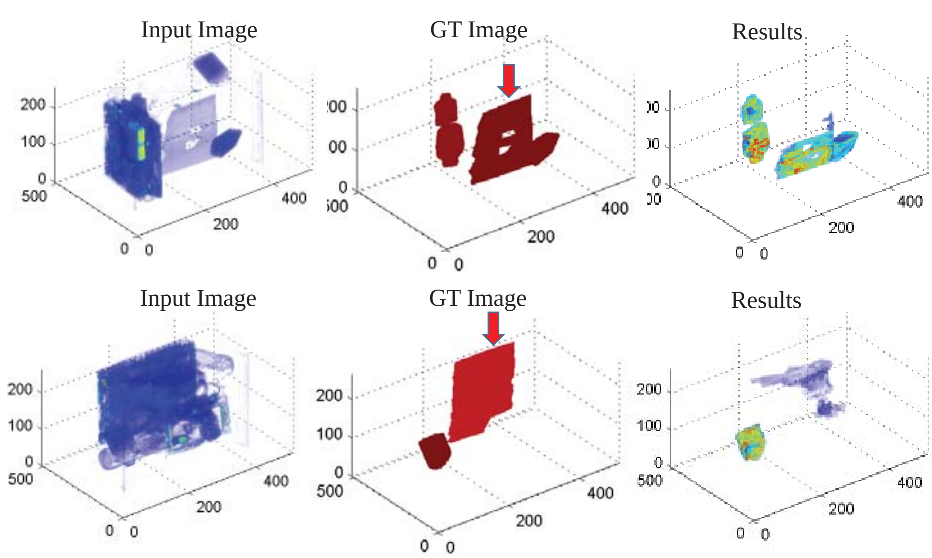




This presentation case, Pseudo-target powder in the plastic bottle, shows the way how the proposed algorithm works with texture in the target area/volume.
➔ The current version of algorithm don't consider these target characteristics (texture) but this can be included in the future versions or could be overcome by the dual energy data (i.e., 2D histogram).

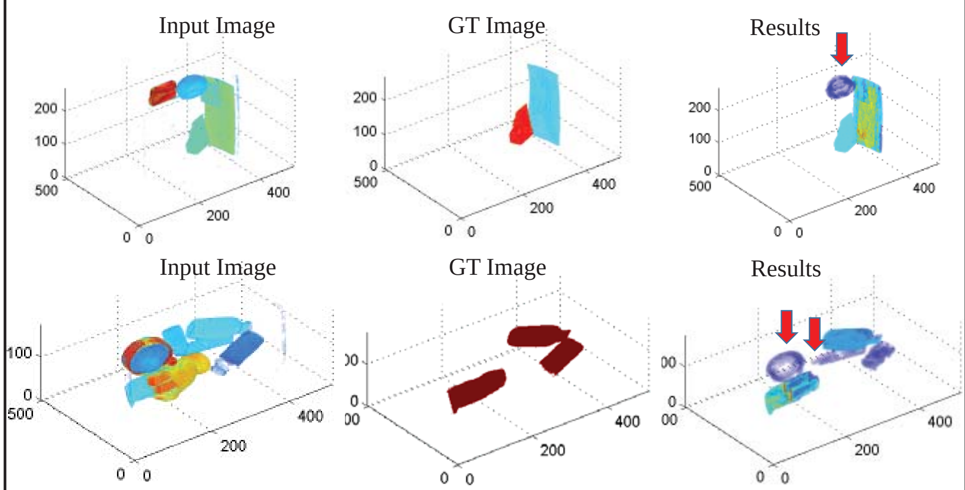
43

Other examples (1/2): Target Missed Cases



44

Other examples (2/2): False Alarm Cases



45

Discussion (1/2)

Strengths:

- It is easy to tune the algorithm for specific targets, which have high priority to detect.
  - Dense material detection, Explosive material detection, Organic material detection, and Liquid detection etc.
- No complicated algorithm training is necessary ➔ An accurate look-up table generation and update would be necessary.
- Easy to implement in the parallel computational architectures.
- It is possible to tune for an individual target performance without changing its overall system performance (by adjusting parameters only for the specific target).
- Easy to extend to multi-spectral data sets ➔ The performance of the proposed algorithm will be improved by many folds (i.e., The core extraction will be executed in two-dimensional space for dual energy scans. And we could use narrow HUBW for each energy data resulting in better sensitivity).

Weaknesses:

- Accurate Histogram measurement or modeling is necessary:
  ➔ Repeated scans for the real targets might be helpful to generate an accurate parameter Look-up table
  ➔ Proposed solution: Histogram sharpening method can be used for real data to improve PD/PFA
- Ground truth image quality and mask generation process are important to the quality of the Look-up table (i.e., noise, blurring, and artifacts could hamper to measure accurate histogram peaks).
  ➔ Potential solution:
  - Advanced image processing methods can be used for a pre-processing (i.e., Denoising, Deburring, MAR, inpainting etc).
  - Advanced image reconstruction method can be used to reduce artifacts and to improve image quality (i.e., Iterative image reconstruction algorithms).

Discussion (2/2)

How this algorithm could be improved in the future.

- Multi-spectral imaging data could improve the performance of the proposed ATR algorithm (i.e., 2D histogram peak locations for Dual energy CT).
- Advanced Image reconstruction method can improve PD/PFA (i.e., iterative image reconstruction algorithm).
- Images with less artifacts could improve overall PD/PFA (i.e., Metal artifacts, Noise, and Blurring etc) ➔ Warping based Metal artifact forecasting method can be used in the future (Attached in the backup slides).

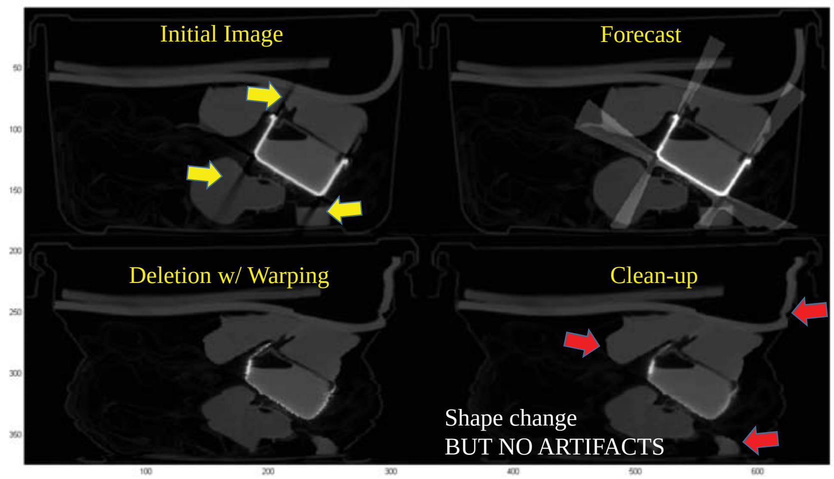
Comments:

- The experimental setups (i.e., target labeling, packing preparation, scanning cases, and data organization etc.) were ideal for new ATR algorithm development.
- The scoring tools were efficient to process multiple cases. But the detected target should be at the same location to get the PD score so that the warping based method couldn't be implemented in this task.
- ATR was a challenging task and still is, there were multiple approaches that were discussed and tested. Through this process in the future, we could not only improve current version of algorithms and theories but contribute to the academia and the security industries.
- Thank you for the support.

Backup Slides

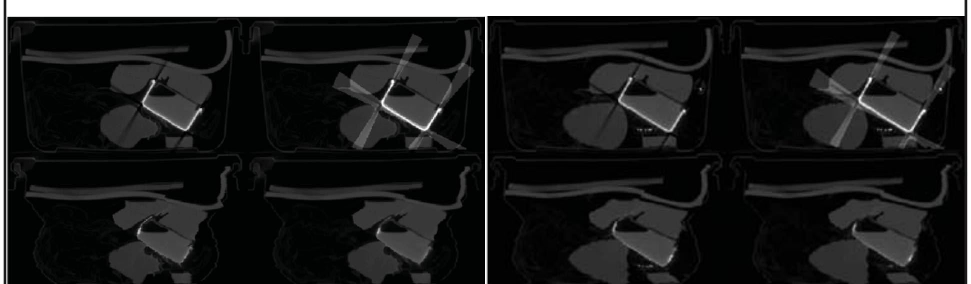
48

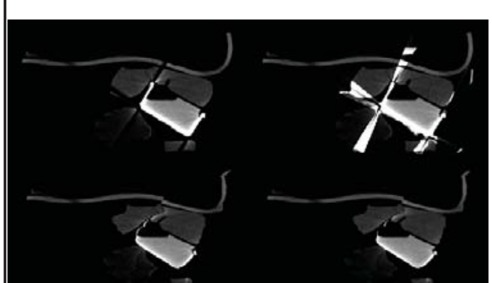
Warping Based Metal Artifact Region Forecast and Removal



A Monkey became a Duck

Performance evaluation



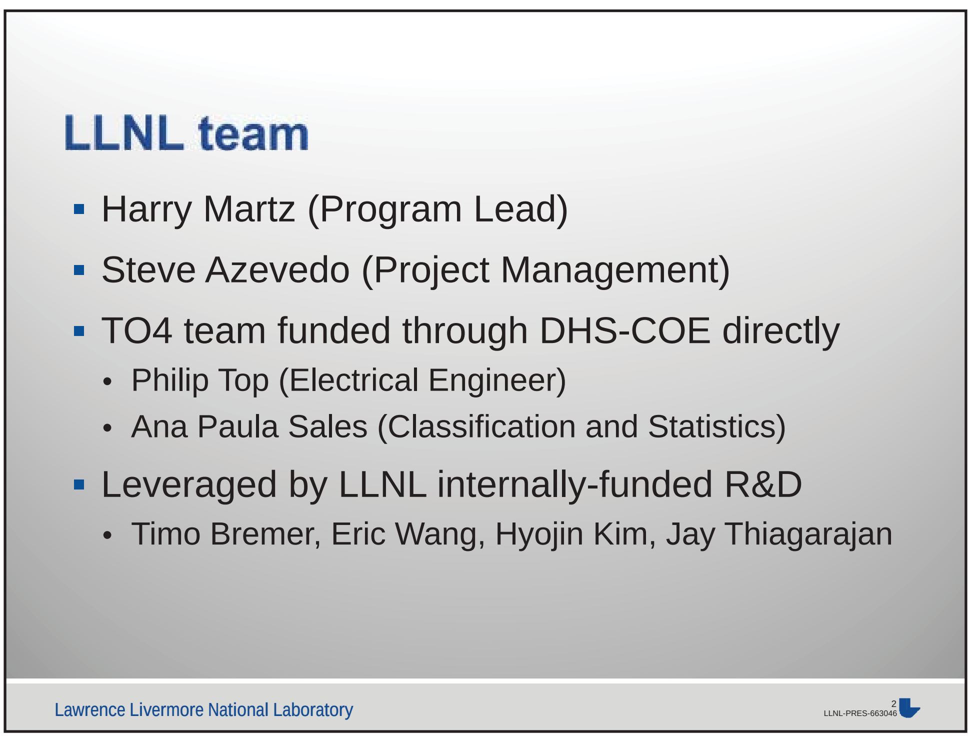


[1000 1800]

Segmentation results are improved. (CT number based)

11.5.2.7 ATR Development – Top

"Automatic Threat Recognition at LLNL"



Summary of P_D/P_{FA} Results

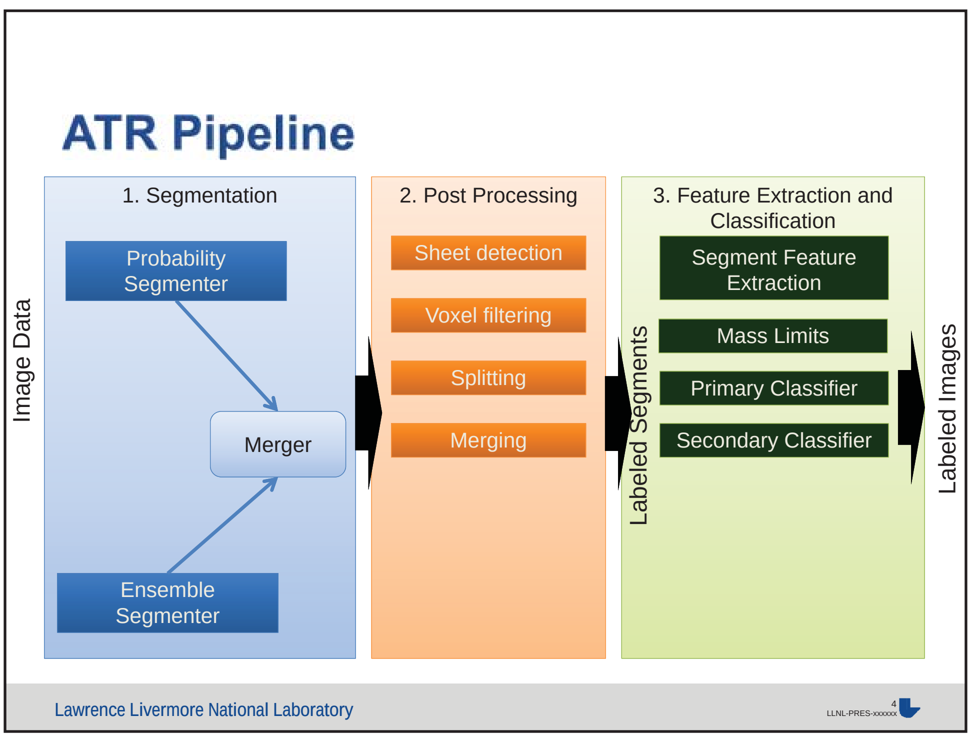
				No special rules (except for PT sheets)	
Target Type	Target Subtype	Level of Difficulty	Num Targets	Num Detected	PD [%]
Target	All	All	407	381	**93.6**
Target	Clay	All	111	107	96.4
Target	Rubber	All	158	150	94.9
Target	Saline	All	138	124	89.9
Target	Bulk	All	270	251	93
Target	Sheet	All	137	130	94.9
Target	All	Low	77	75	97.4
Target	Clay	Low	29	29	100
Target	Rubber	Low	22	22	100
Target	Saline	Low	26	24	92.3
Target	Bulk	Low	56	54	96.4
Target	Sheet	Low	21	21	100
Target	All	High	317	294	92.7
Target	Clay	High	82	78	95.1
Target	Rubber	High	125	118	94.4
Target	Saline	High	110	98	89.1
Target	Bulk	High	201	185	92
Target	Sheet	High	116	109	94
Pseudo-target	Sheet	High	10	10	**100**
			Num Non-targets	Num FAs	PFA [%]
			1371	163	**11.9**
				Num Scans with FAs	Avg Num FAs
				110	1.57

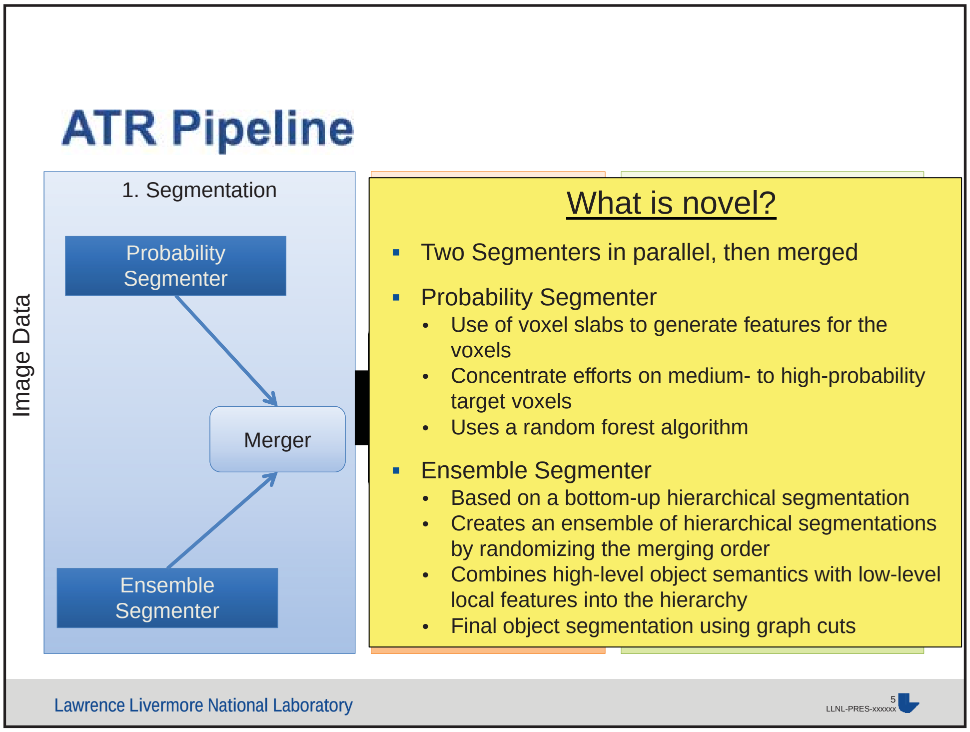
No special rules:
93.6% / 11.9%

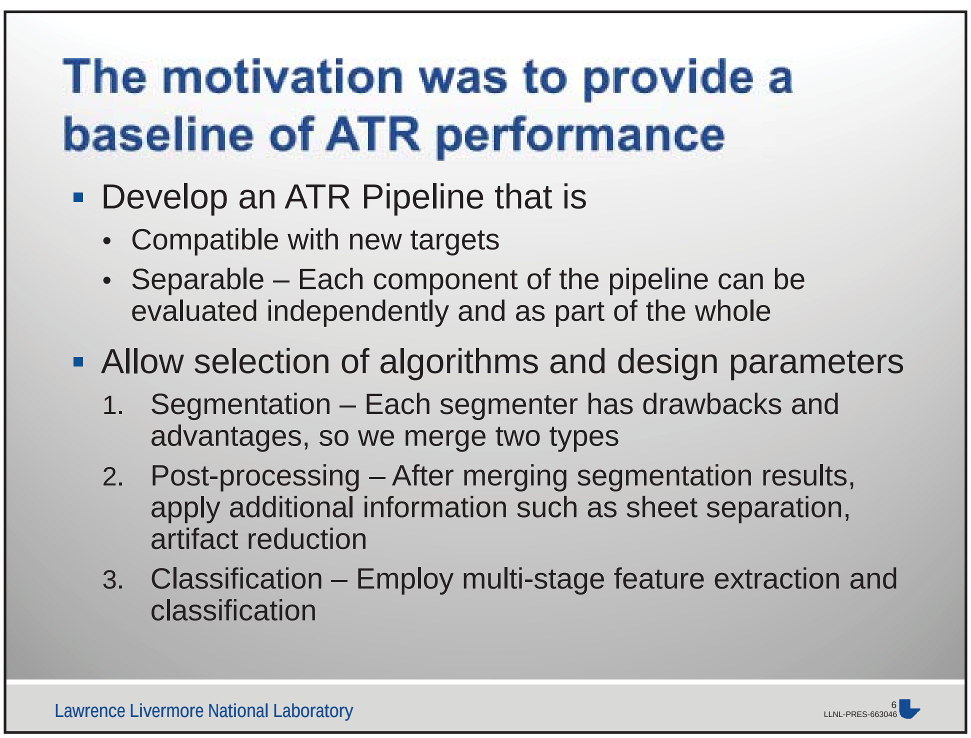
(special post-processing for pseudo-target sheets)

Lawrence Livermore National Laboratory

LLNL-PRES-663046 3

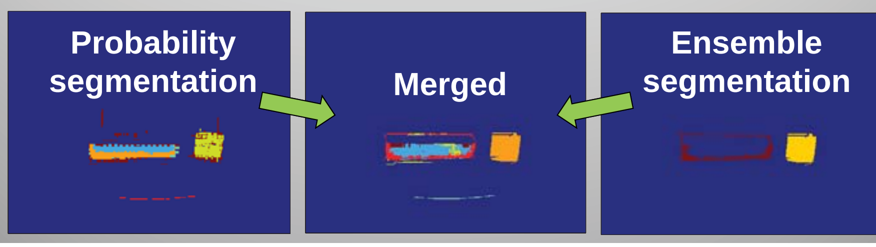



ATR Pipeline
Image Data
1. Segmentation
Probability Segmenter
Merger
Ensemble Segmenter
What is novel?
▪ Two Segmenters in parallel, then merged
▪ Probability Segmenter
• Use of voxel slabs to generate features for the voxels
• Concentrate efforts on medium- to high-probability target voxels
• Uses a random forest algorithm
▪ Ensemble Segmenter
• Based on a bottom-up hierarchical segmentation
• Creates an ensemble of hierarchical segmentations by randomizing the merging order
• Combines high-level object semantics with low-level local features into the hierarchy
• Final object segmentation using graph cuts
Lawrence Livermore National Laboratory
LLNL-PRES-xxxxxx
5


The motivation was to provide a baseline of ATR performance
▪ Develop an ATR Pipeline that is
• Compatible with new targets
• Separable – Each component of the pipeline can be evaluated independently and as part of the whole
▪ Allow selection of algorithms and design parameters
1. Segmentation – Each segmenter has drawbacks and advantages, so we merge two types
2. Post-processing – After merging segmentation results, apply additional information such as sheet separation, artifact reduction
3. Classification – Employ multi-stage feature extraction and classification
Lawrence Livermore National Laboratory
LLNL-PRES-663046
6

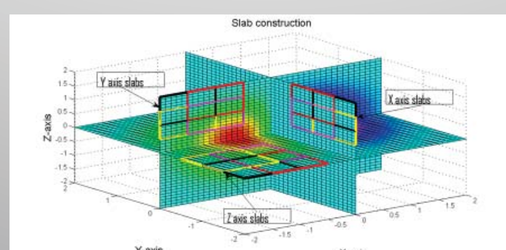
1. Segmentation

- Probability segmenter
  - Compute a probability that each voxel belongs to a target and merge connected voxels together
  - Goal of 100% recall
  - Tends to merge targets together
  - Poor precision
- Ensemble segmenter
  - Generate an ensemble of potential segments and compute based on average behavior
  - Good segmentation
  - Goal of >90% detection
  - It misses several sheet objects



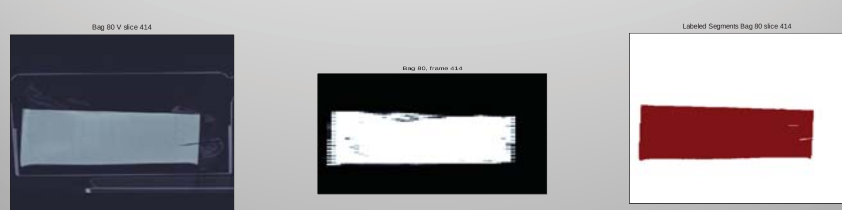
Technical Description of Algorithm: Probability Segmenter (1 of 2)

- Break image into 10x10 voxel "slabs" in each plane X,Y,Z (planes, not cubes)
- Generate a feature vector from the slab
  - median, stdev, range, type-dependent features based on the discrete cosine transform for texture
- Compute the probability the slab belongs to a target of interest (clay, saline, rubber, and powder), or below a threshold



Technical Description of Algorithm: Probability Segmenter (2 of 2)

- Subject the identified slabs to 3D connected-component labeling
  - Only slabs that are connected (via adjacency) to at least K other slabs (to form segments of a minimum size) are retained
  - Slabs that are not connected to enough other slabs are discarded
- The output of the Probability Segmenter are these "rough segments"



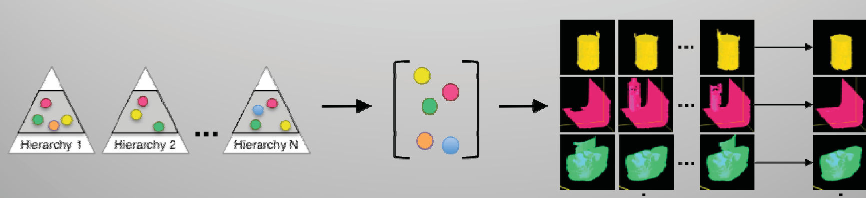
Bag 80 y-slice 414 | Compute probabilities | Threshold and find connected voxels

Lawrence Livermore National Laboratory

LLNL-PRES-663046 9

Technical Description of Algorithm: Ensemble Segmenter (1 of 2)

- Creates an ensemble of hierarchical segmentations by randomizing the merging order of local features (attenuation, histogram)
- Include high-level object semantics (e.g., surface/volume ratio) with low-level local features into hierarchy of candidate objects
- Combine localized candidate objects into final objects using consensus segmentation with graphcuts



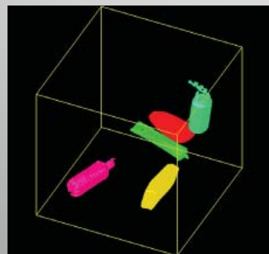
Ensemble of hierarchical segmentations | Set of all candidate objects | Localized candidate objects | Final objects

Lawrence Livermore National Laboratory

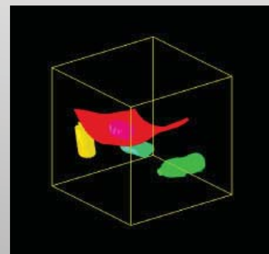
LLNL-PRES-663046 10

Technical Description of Algorithm: Ensemble Segmenter (2 of 2)

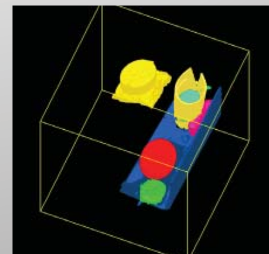
- For many object types, it compensates for reconstruction artifacts
- Objects can be identified from a wide range of levels in the hierarchy
- Can be customized for how much segmentation is desired
- Converges to the "average" behavior with consensus segmentation



SSN: 088



SSN: 093



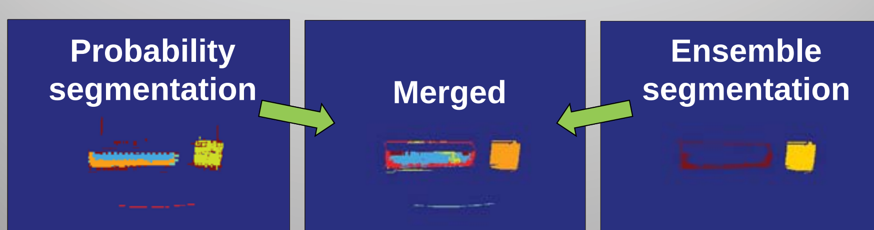
SSN: 094

Lawrence Livermore National Laboratory

11 LLNL-PRES-663046

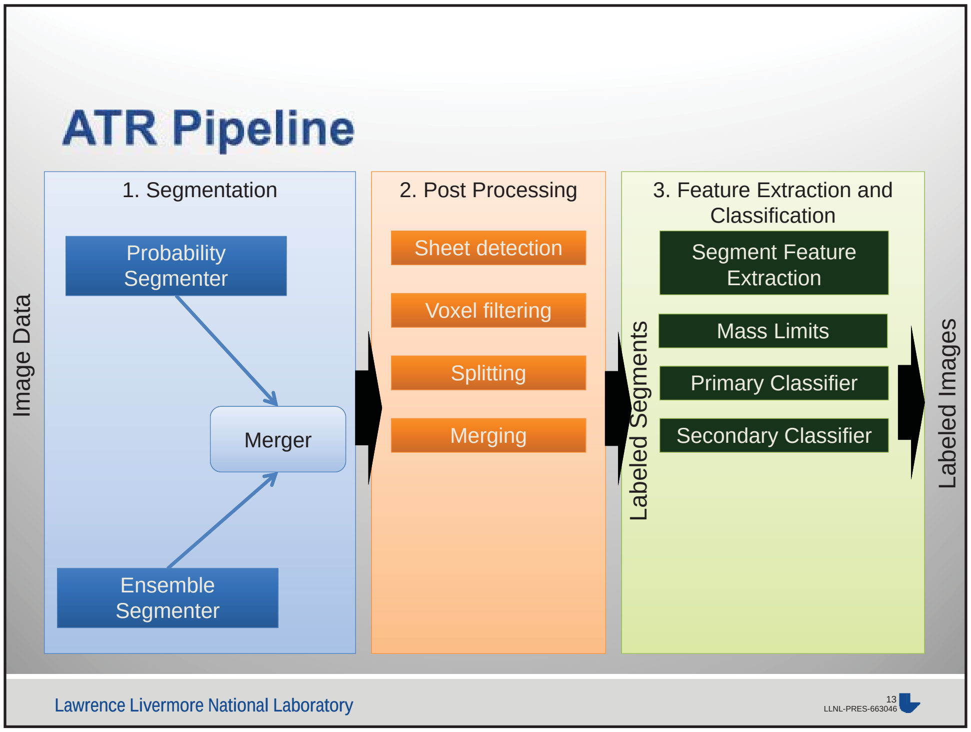
1. Flexible Segmentation Merger

- Use segments from Ensemble Segmenter unless very few pixels are found in the Probability Segmenter (# pixels is flexible)
- Make new segments out of the remaining Probability Segmenter voxels
- A segment can be split if only half is in Probability Segmenter
- Complexity can be traded off with the post-processing



Lawrence Livermore National Laboratory

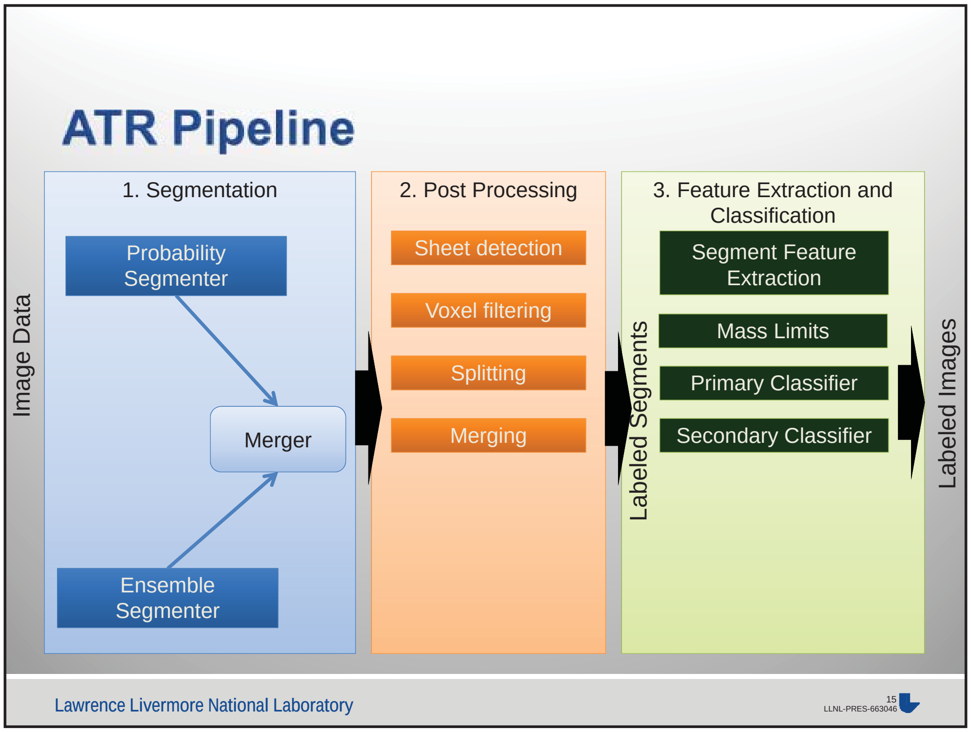
12 LLNL-PRES-663046

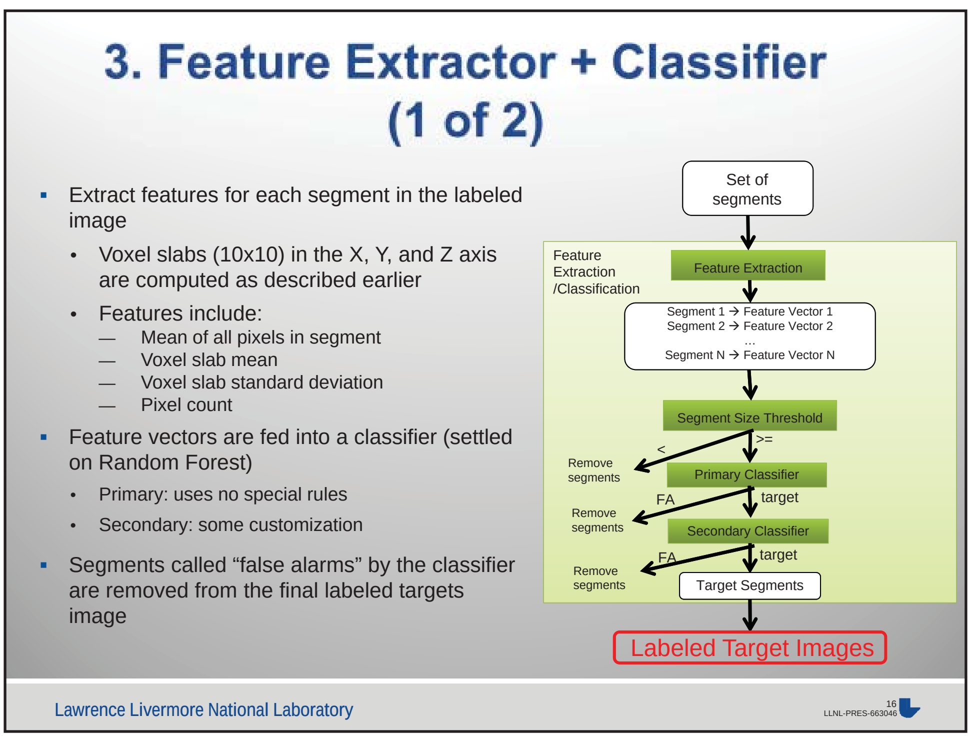


2. Post-processing

- Performs further splits of segments
  - Are large sections of segments only connected by a narrow channel or not connected? (If so, split.) OR
  - Are there multiple statistically separable histogram peaks? (If so, split.) OR
  - Can one of the adjoining segments be characterized as a sheet? (If so, remove any large clusters.)
- Performs further merges of segments
  - Are segments close together or overlapping? AND
  - Do they have the same statistical properties? AND
  - Were they previously separated? AND
  - Do they fit together?
  - (If so, merge them.)

Lawrence Livermore National Laboratory
LLNL-PRES-663046
14


ATR Pipeline
Image Data
1. Segmentation
Probability Segmenter
Merger
Ensemble Segmenter
2. Post Processing
Sheet detection
Voxel filtering
Splitting
Merging
Labeled Segments
3. Feature Extraction and Classification
Segment Feature Extraction
Mass Limits
Primary Classifier
Secondary Classifier
Labeled Images
Lawrence Livermore National Laboratory
15
LLNL-PRES-663046


3. Feature Extractor + Classifier
(1 of 2)
▪ Extract features for each segment in the labeled image
• Voxel slabs (10x10) in the X, Y, and Z axis are computed as described earlier
• Features include:
— Mean of all pixels in segment
— Voxel slab mean
— Voxel slab standard deviation
— Pixel count
▪ Feature vectors are fed into a classifier (settled on Random Forest)
• Primary: uses no special rules
• Secondary: some customization
▪ Segments called "false alarms" by the classifier are removed from the final labeled targets image
Set of segments
Feature Extraction /Classification
Feature Extraction
Segment 1 → Feature Vector 1
Segment 2 → Feature Vector 2
...
Segment N → Feature Vector N
Segment Size Threshold
<
>=
Remove segments
Primary Classifier
FA
target
Remove segments
Secondary Classifier
FA
target
Remove segments
Target Segments
Labeled Target Images
Lawrence Livermore National Laboratory
16
LLNL-PRES-663046

3. Feature Extractor + Classifier (2 of 2)

- Feature Extractor
  - Segment size threshold: segments smaller than a certain threshold are removed (i.e., labeled as non-target)
  - Pseudo-target sheet threshold: segments that are within a certain range for a number of features are retained in the final labeled images
    - Examples of features are mean, mode, and standard deviation of attenuation, number of peaks, cosine transform
- Primary Classifier – Operates on the entire set
  - Used a Random Forest (RF) algorithm
  - We use 3 RFs: train on 1/3 of the data, and evaluate the other 2/3; for each 1/3 of data
- Secondary Classifier – Allows other rules to reduce FA
  - Provides further filtering for the segments that pass the Primary Classifier
  - Particular rules for pseudo-target sheets
  - Also based on RFs, but training and evaluation sets differ somewhat

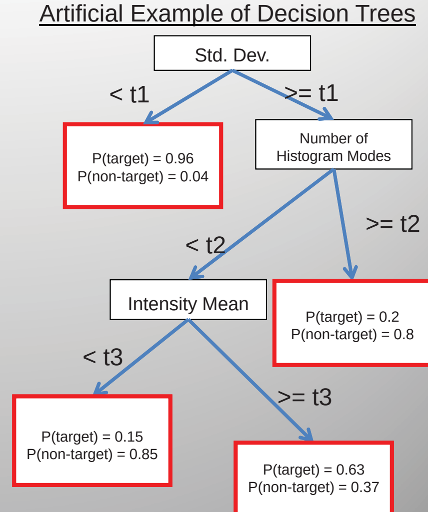
Lawrence Livermore National Laboratory

17
LLNL-PRES-663046

Random Forests Classifiers

- A Random Forest is an ensemble of decision trees
  - Features are selected at features
- Decision trees
  - Provide a partitioning of the feature space of the data into disjoint sets
  - Each partitioning is associated with a probability vector of the possible outcome classes
  - Classification of a new object is done by mapping features to the partitioning of the data
  - The label is defined by the probability vector

Artificial Example of Decision Trees



Lawrence Livermore National Laboratory

18
LLNL-PRES-663046

Other Classifiers were explored

(that did not work as well as Random Forests)

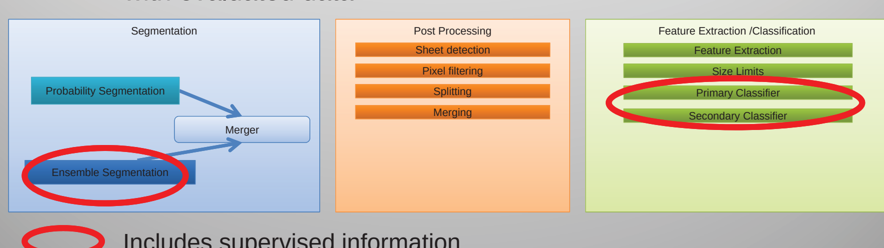
- Adaptive Boosting
  - — An ensemble classifier where the outputs of classifiers are weighted according to how weak/strong they are. Weak classifiers are tweaked in favor of those instances misclassified by previous classifiers
- Artificial Neural Networks
  - — Algorithm inspired by how information is transmitted in the brain via neurons. Large number of inputs are approximated by layers of neurons whose connections are learned.
- Naïve Bayes
  - — Probabilistic classifier based on Bayes theorem. It assumes independence of features.
- Nearest Neighbors
  - — Provides simple data interpolation in one or many dimensions. It clusters the training data, and each cluster is represented by its centroid. New observations are assigned to the cluster whose centroid is most similar to itself.
- Support Vector Machines
  - — Obtains data classification by identifying an optimal hyperplane that separates the two classes under consideration.

Lawrence Livermore National Laboratory

19
LLNL-PRES-663046

ATR Training

- How was over-training on supplied data prevented?
  - Majority of steps are unsupervised – no special rules
  - Supervised steps use three-fold cross-validation (1/3 training, 2/3 evaluation)
    - — Use multiple classifiers such that training data never overlaps with evaluated data



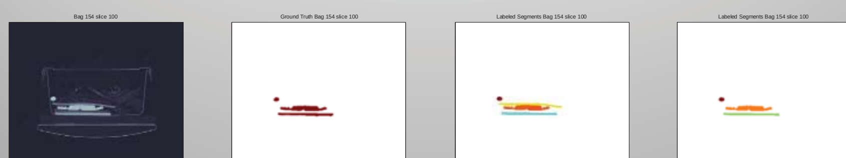
Includes supervised information

Lawrence Livermore National Laboratory

20
LLNL-PRES-663046

ATR Training

- How were false alarms reduced?
  - By the use of multiple staged steps
    - Probability Segmenter: Reduces the number of voxels used in segmentation. It is tuned to have nearly 100% recall and minimize the number of false alarms.
    - Classifier: Labels the segments as “targets” or “false alarms”, such that only the “target” segments are included in the final set of labeled images.



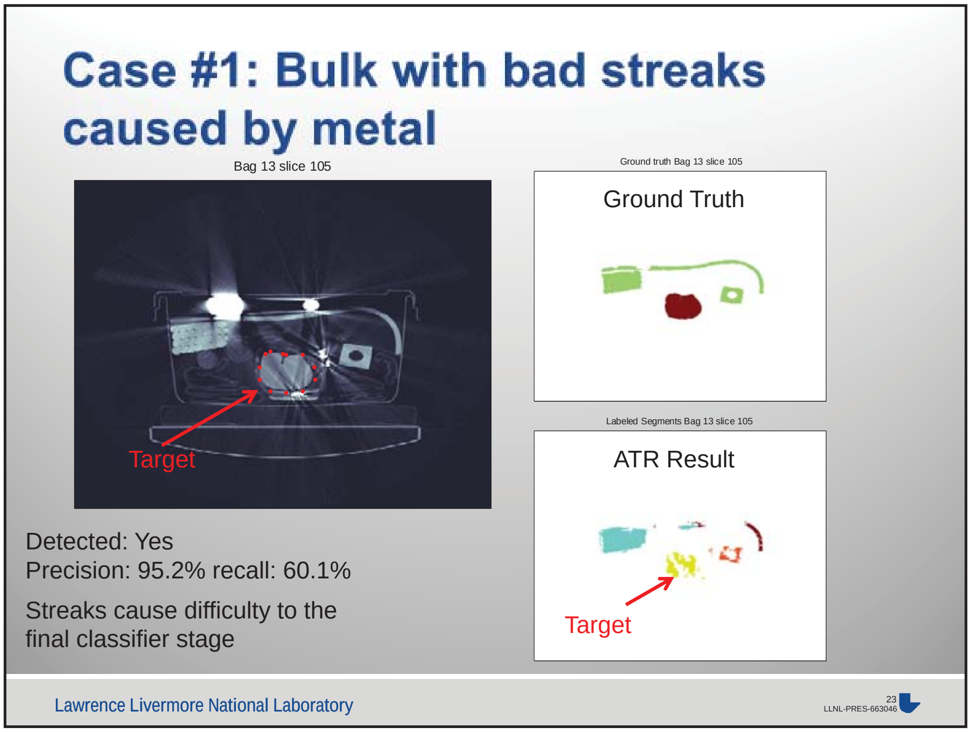
CT | Ground Truth | After Preprocessing | Final

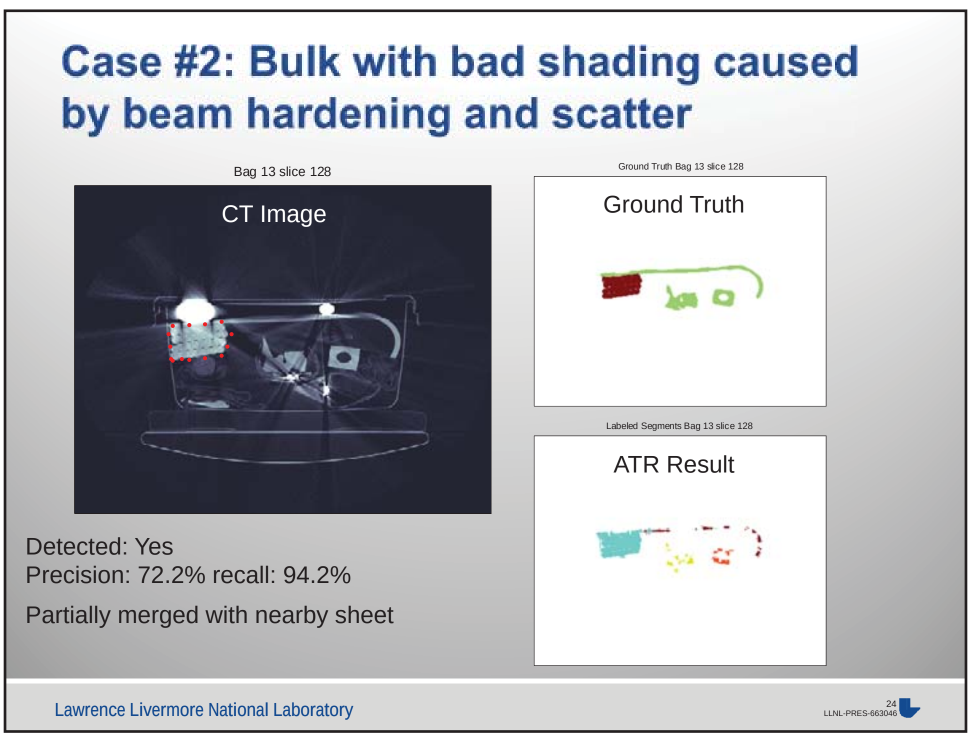
Lawrence Livermore National Laboratory
21
LLNL-PRES-663046

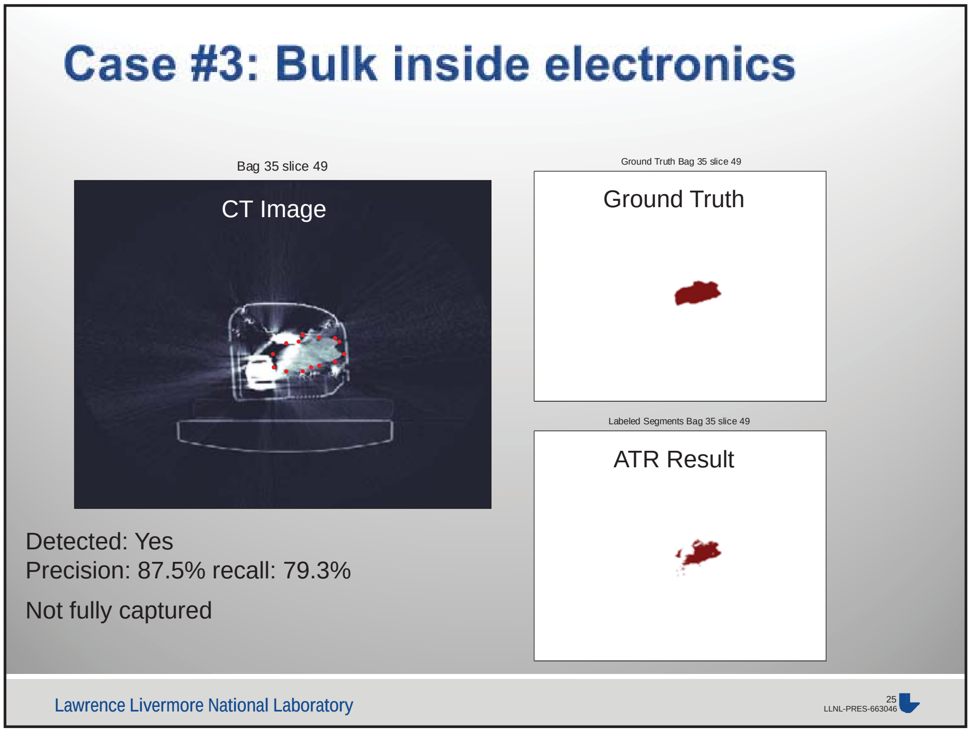
Robustness to new targets

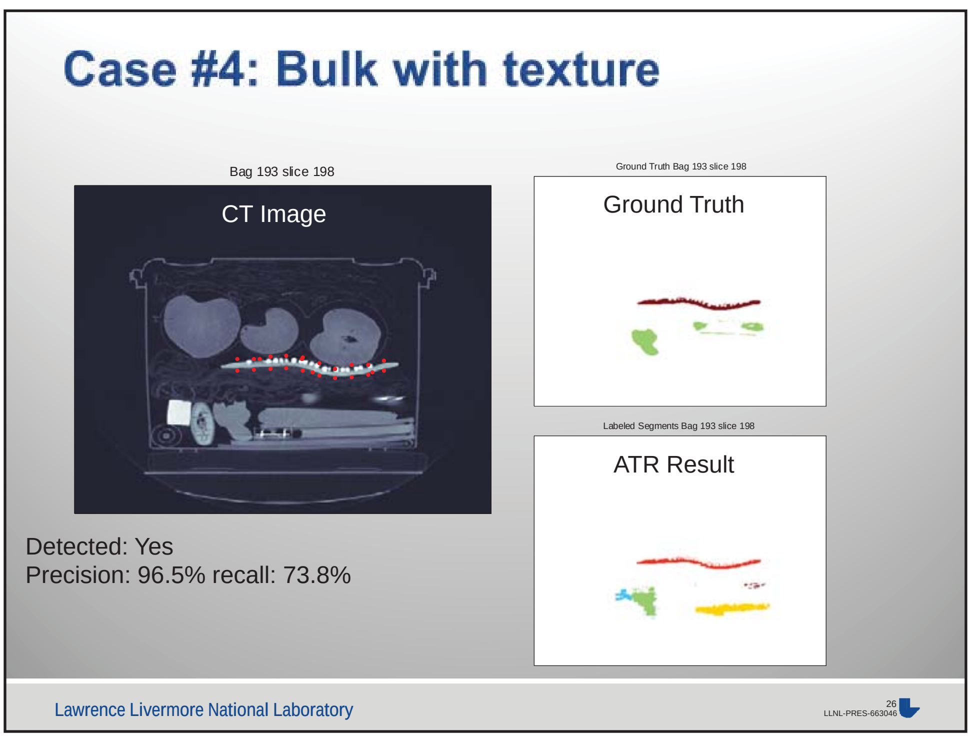
- Each of the target types has a separate path in the pipeline… starting from segmentation
- This facilitates the addition or removal of targets
- We can use simulated data to detect new target classes; all that is needed are the features

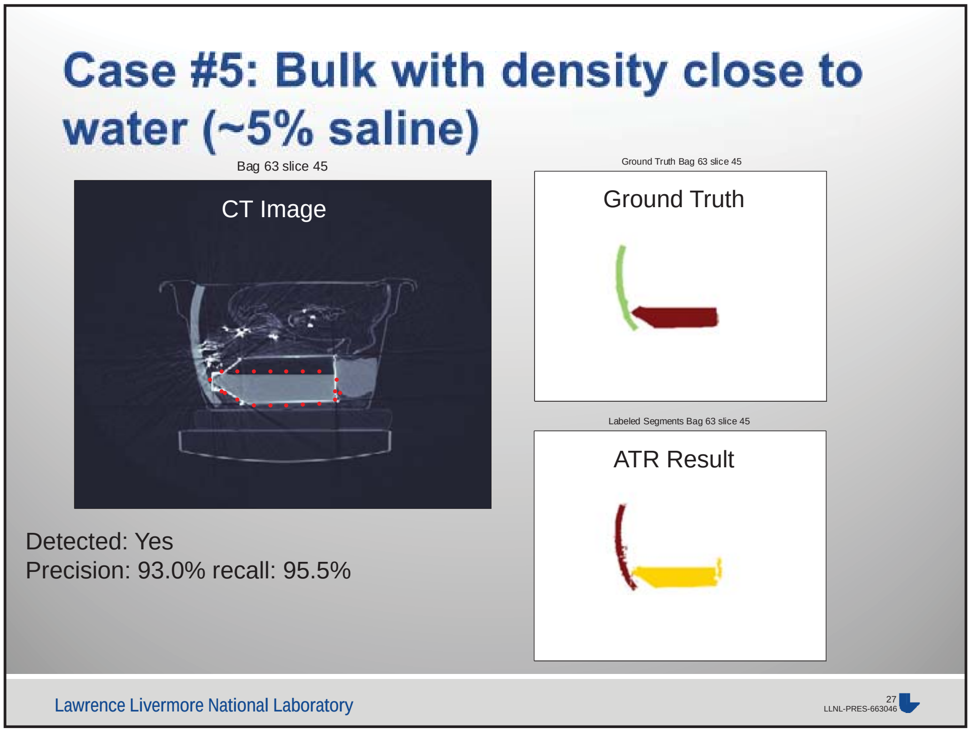
Lawrence Livermore National Laboratory
22
LLNL-PRES-663046

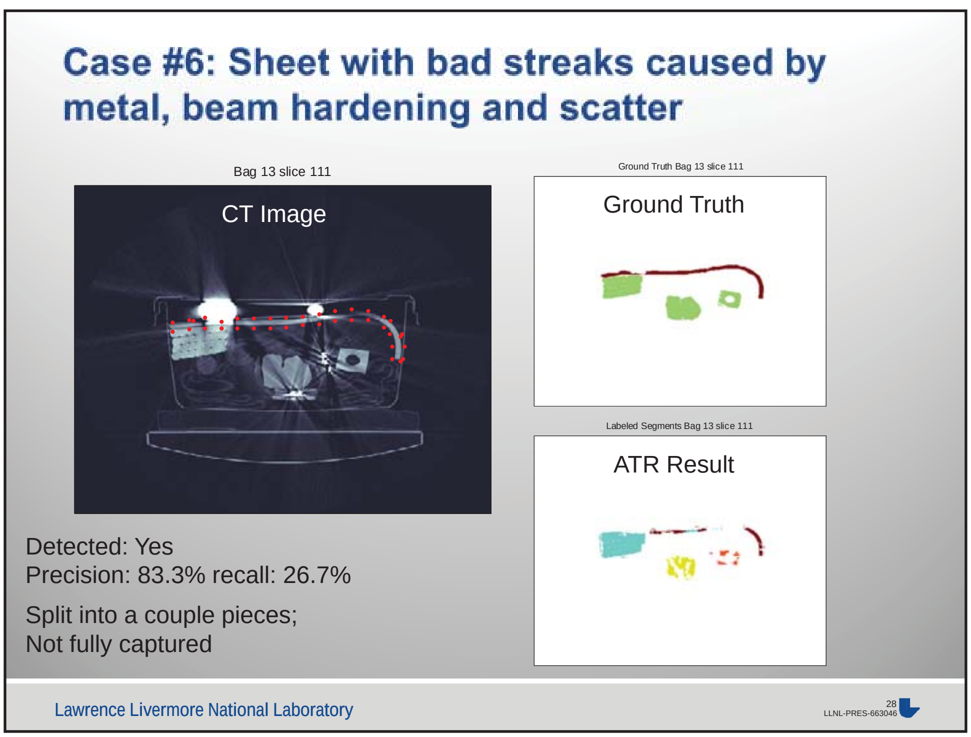

Case #1: Bulk with bad streaks caused by metal
Bag 13 slice 105
Target
Ground truth Bag 13 slice 105
Ground Truth
Labeled Segments Bag 13 slice 105
ATR Result
Target
Detected: Yes
Precision: 95.2% recall: 60.1%
Streaks cause difficulty to the final classifier stage
Lawrence Livermore National Laboratory
23
LLNL-PRES-663046

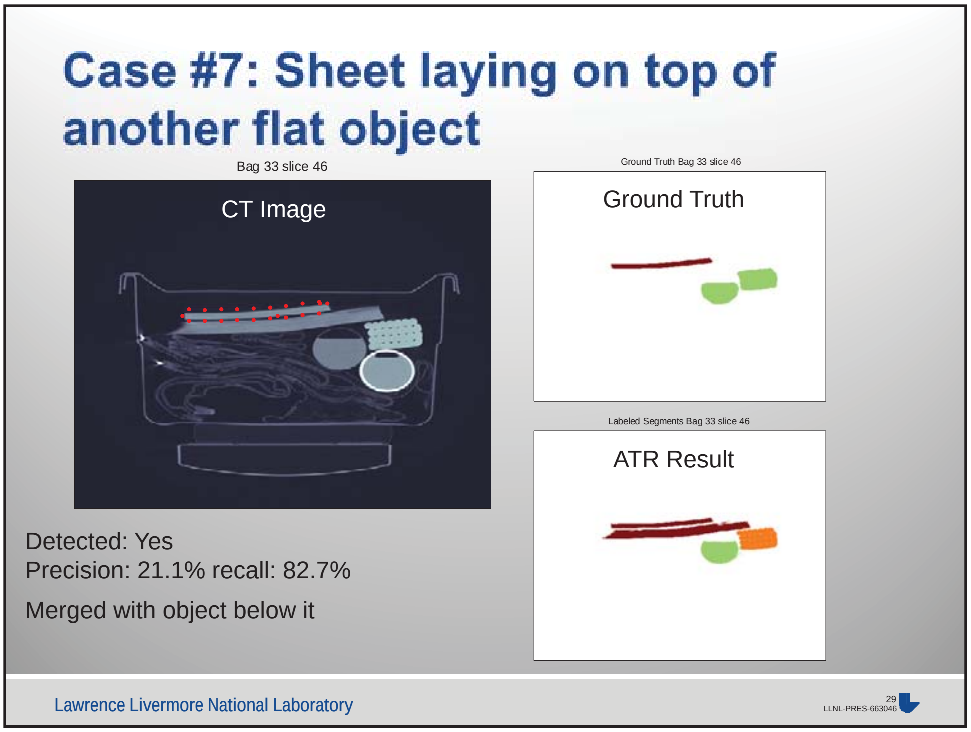

Case #2: Bulk with bad shading caused by beam hardening and scatter
Bag 13 slice 128
CT Image
Ground Truth Bag 13 slice 128
Ground Truth
Labeled Segments Bag 13 slice 128
ATR Result
Detected: Yes
Precision: 72.2% recall: 94.2%
Partially merged with nearby sheet
Lawrence Livermore National Laboratory
24
LLNL-PRES-663046

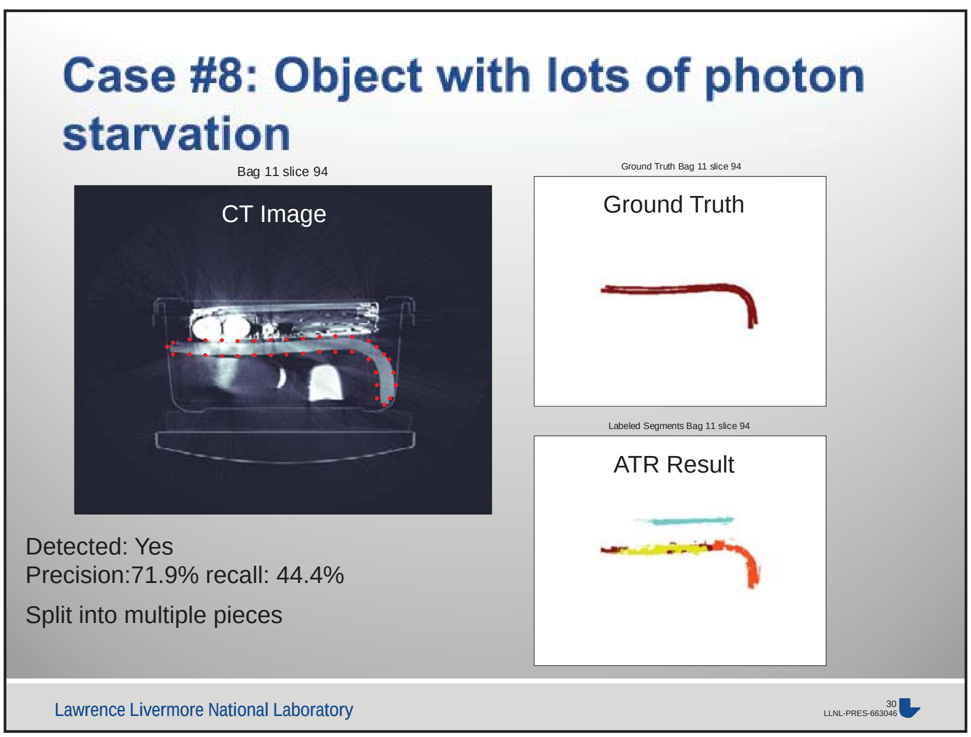

Case #3: Bulk inside electronics
Bag 35 slice 49
CT Image
Ground Truth Bag 35 slice 49
Ground Truth
Labeled Segments Bag 35 slice 49
ATR Result
Detected: Yes
Precision: 87.5% recall: 79.3%
Not fully captured
Lawrence Livermore National Laboratory
25
LLNL-PRES-663046


Case #4: Bulk with texture
Bag 193 slice 198
CT Image
Ground Truth Bag 193 slice 198
Ground Truth
Labeled Segments Bag 193 slice 198
ATR Result
Detected: Yes
Precision: 96.5% recall: 73.8%
Lawrence Livermore National Laboratory
26
LLNL-PRES-663046


Case #5: Bulk with density close to water (~5% saline)
Bag 63 slice 45
CT Image
Ground Truth Bag 63 slice 45
Ground Truth
Labeled Segments Bag 63 slice 45
ATR Result
Detected: Yes
Precision: 93.0% recall: 95.5%
Lawrence Livermore National Laboratory
27
LLNL-PRES-663046

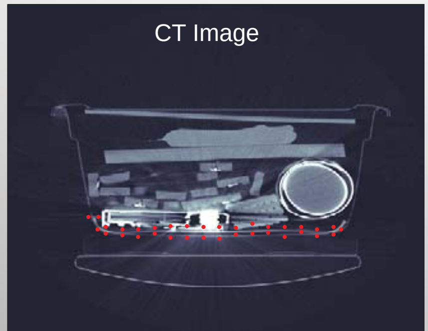

Case #6: Sheet with bad streaks caused by metal, beam hardening and scatter
Bag 13 slice 111
CT Image
Ground Truth Bag 13 slice 111
Ground Truth
Labeled Segments Bag 13 slice 111
ATR Result
Detected: Yes
Precision: 83.3% recall: 26.7%
Split into a couple pieces;
Not fully captured
Lawrence Livermore National Laboratory
28
LLNL-PRES-663046


Case #7: Sheet laying on top of another flat object
Bag 33 slice 46
CT Image
Ground Truth Bag 33 slice 46
Ground Truth
Labeled Segments Bag 33 slice 46
ATR Result
Detected: Yes
Precision: 21.1% recall: 82.7%
Merged with object below it
Lawrence Livermore National Laboratory
29
LLNL-PRES-663046

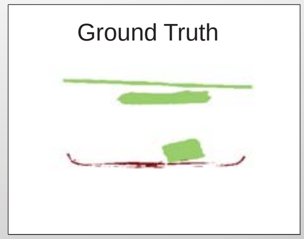

Case #8: Object with lots of photon starvation
Bag 11 slice 94
CT Image
Ground Truth Bag 11 slice 94
Ground Truth
Labeled Segments Bag 11 slice 94
ATR Result
Detected: Yes
Precision:71.9% recall: 44.4%
Split into multiple pieces
Lawrence Livermore National Laboratory
30
LLNL-PRES-663046

Case #9: PT sheet based on thickness

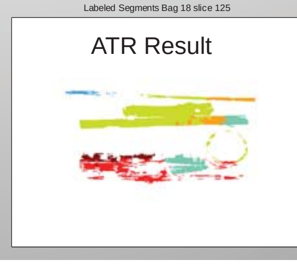
Bag 18 slice 125



Ground Truth Bag 18 slice 125



Labeled Segments Bag 18 slice 125



Detected: Yes
Precision: 23.2% recall: 32.6%

Not well captured and merged with some surroundings

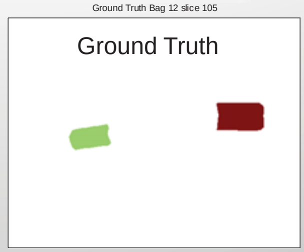
Lawrence Livermore National Laboratory

31
LLNL-PRES-663046

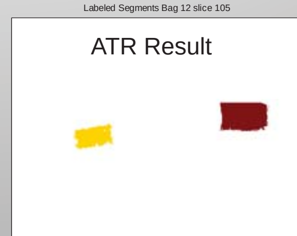
Case #10: PT Powder (based on density, not mass)

Bag 12 slice 105

Ground Truth Bag 12 slice 105



Labeled Segments Bag 12 slice 105

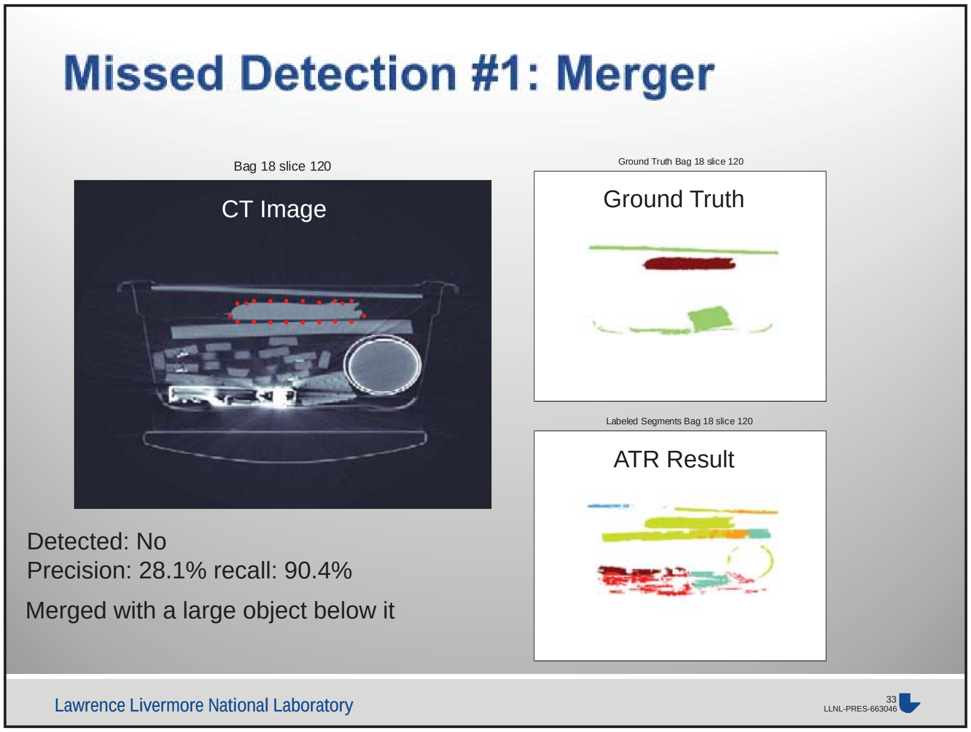


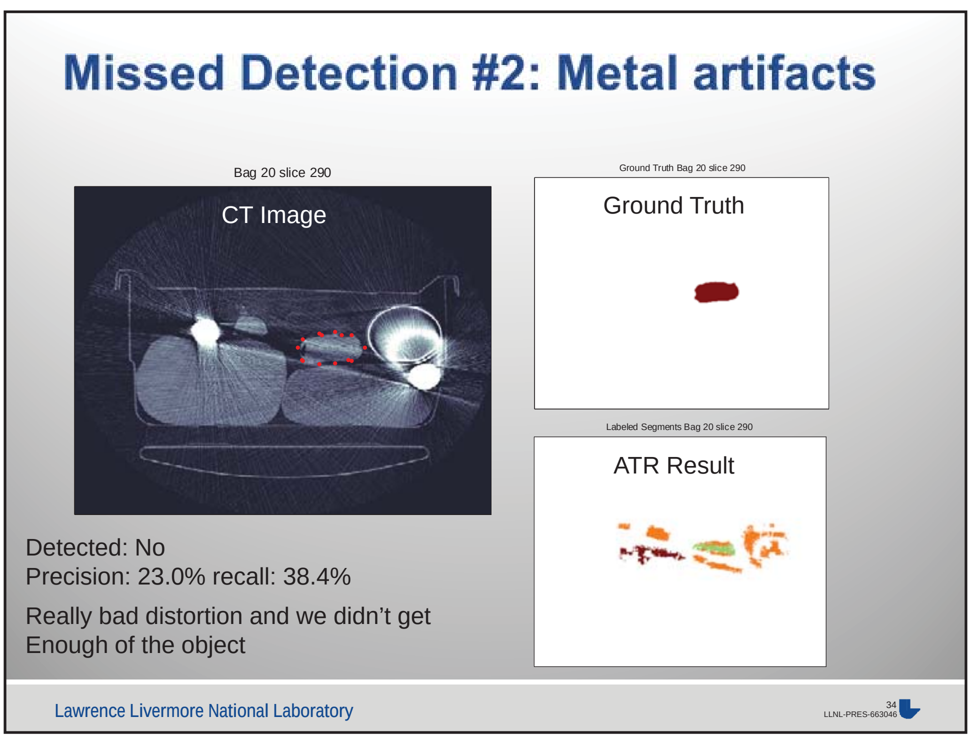
Detected: No
Precision: 49.95% recall: 96.0%

Merged with another object (behind in 3D)

Lawrence Livermore National Laboratory

32
LLNL-PRES-663046


Missed Detection #1: Merger
Bag 18 slice 120
CT Image
Ground Truth Bag 18 slice 120
Ground Truth
Labeled Segments Bag 18 slice 120
ATR Result
Detected: No
Precision: 28.1% recall: 90.4%
Merged with a large object below it
Lawrence Livermore National Laboratory
33
LLNL-PRES-663046


Missed Detection #2: Metal artifacts
Bag 20 slice 290
CT Image
Ground Truth Bag 20 slice 290
Ground Truth
Labeled Segments Bag 20 slice 290
ATR Result
Detected: No
Precision: 23.0% recall: 38.4%
Really bad distortion and we didn't get
Enough of the object
Lawrence Livermore National Laboratory
34
LLNL-PRES-663046

Discussion

- Strengths – flexibility, robustness to new object characteristics and types
- Weakness – misses some of the targets
  - 7 missed because merged with another object
  - 4 missed due to splitting
  - Bag 18 and Bag 20 split and merged

Lawrence Livermore National Laboratory

LLNL-PRES-663046 35

13 Missed Targets

- 4 were split, 7 were merged with another object
- Bag 18 was split and merged with multiple objects
- Bag 20 was split and merged with a false alarm object

Bag	Target	type	Prec	recall
13	6047	R	92	47
15*	6045	C	98	46
16*	6002	S	33	97
18*	6025	S	28	90
18*	6051	C	79	32
18*	8031	R sh	5	17
20	6012	S	23	38

Bag	Target	type	Prec	recall
34	6012	S	95	43
38	6001	S	41	98
115	6178	S	46	92
147*	6140	R sh	18	65
162*	6573	R sh	15	93
183	6557	S	20	65

* Object detected in some tests but not current best results

Lawrence Livermore National Laboratory

LLNL-PRES-663046 36

Future improvements

- Improve the merge/split algorithms
- Improve sheet processing
- Better final stage classifiers
- Better image reconstruction to reduce streaking artifacts (out of scope for this project), e.g., MAR
- Code refinements
- Apply algorithms to other data sets including potentially classified systems

Lawrence Livermore National Laboratory

37
LLNL-PRES-663046

Comments on the data

- The definition of false alarms creates potentially misleading precision and recall results.
  - A single target split into two parts both of which are detected this situation creates 2 false alarm objects and 1 missed detection
  - Two targets merged together likewise creates 1 false alarm and 2 missed detections

Lawrence Livermore National Laboratory

38
LLNL-PRES-663046

What you learned by participating in this process

- Segmentation is the heart of this problem.
- Algorithm tuning is heavily dependent on the rules of the test
  - We are forced to "tune to the test", which makes the final result less robust to new data whether targets or not. A blind test would be ideal.
- ATR of luggage from CT is hard
  - There is an overlap of targets and non-targets
  - Physics alone is insufficient to get perfection
  - Overtraining is sometimes needed to pass a test – and can lead to 100/0 performance for any known set

Summary of P_D/P_{FA} Results

				No special rules (except for PT sheets)	
Target Type	Target Subtype	Level of Difficulty	Num Targets	Num Detected	PD [%]
Target	All	All	407	381	**93.6**
Target	Clay	All	111	107	96.4
Target	Rubber	All	158	150	94.9
Target	Saline	All	138	124	89.9
Target	Bulk	All	270	251	93
Target	Sheet	All	137	130	94.9
Target	All	Low	77	75	97.4
Target	Clay	Low	29	29	100
Target	Rubber	Low	22	22	100
Target	Saline	Low	26	24	92.3
Target	Bulk	Low	56	54	96.4
Target	Sheet	Low	21	21	100
Target	All	High	317	294	92.7
Target	Clay	High	82	78	95.1
Target	Rubber	High	125	118	94.4
Target	Saline	High	110	98	89.1
Target	Bulk	High	201	185	92
Target	Sheet	High	116	109	94
Pseudo-target	Sheet	High	10	10	**100**
			Num Non-targets	Num FAs	PFA [%]
			1371	163	**11.9**
				Num Scans with FAs	Avg Num FAs
				110	1.57

No special rules:
93.6% / 11.9%

Summary of P_D/P_{FA} Results

				No special rules (except for PT sheets)		New rules added for corner cases	
Target Type	Target Subtype	Level of Difficulty	Num Targets	Num Detected	PD [%]	Num Detected	PD [%]
Target	All	All	407	381	**93.6**	387	**95.1**
Target	Clay	All	111	107	96.4	107	96.4
Target	Rubber	All	158	150	94.9	151	95.6
Target	Saline	All	138	124	89.9	129	93.5
Target	Bulk	All	270	251	93	256	94.8
Target	Sheet	All	137	130	94.9	131	95.6
Target	All	Low	77	75	97.4	77	100
Target	Clay	Low	29	29	100	29	100
Target	Rubber	Low	22	22	100	22	100
Target	Saline	Low	26	24	92.3	26	100
Target	Bulk	Low	56	54	96.4	56	100
Target	Sheet	Low	21	21	100	21	100
Target	All	High	317	294	92.7	298	94
Target	Clay	High	82	78	95.1	78	95.1
Target	Rubber	High	125	118	94.4	119	95.2
Target	Saline	High	110	98	89.1	101	91.8
Target	Bulk	High	201	185	92	188	93.5
Target	Sheet	High	116	109	94	110	94.8
Pseudo-target	Sheet	High	10	10	**100**	10	**100**
			Num Non-targets	Num FAs	PFA [%]	Num FAs	PFA [%]
			1371	163	**11.9**	15	**1.1**
				Num Scans with FAs	Avg Num FAs	Num Scans with FAs	Avg Num FAs
				110	1.57	15	1

No special rules:
93.6% / 11.9%

"Over-trained":
95.1% / 1.1%

Lawrence Livermore National Laboratory

41
LLNL-PRES-663046

Summary of P_D/P_{FA} Results

				No special rules (except for PT sheets)		New rules added for corner cases	
Target Type	Target Subtype	Level of Difficulty	Num Targets	Num Detected	PD [%]	Num Detected	PD [%]
Target	All	All	407	381	**93.6**	387	**95.1**
Target	Clay	All	111	107	96.4	107	96.4
Target	Rubber	All	158	150	94.9	151	95.6
Target	Saline	All	138	124	89.9	129	93.5
Target	Bulk	All	270	251	93	256	94.8
Target	Sheet	All	137	130	94.9	131	95.6
Target	All	Low	77	75	97.4	77	100
Target	Clay	Low	29	29	100	29	100
Target	Rubber	Low	22	22	100	22	100
Target	Saline	Low	26	24	92.3	26	100
Target	Bulk	Low	56	54	96.4	56	100
Target	Sheet	Low	21	21	100	21	100
Target	All	High	317	294	92.7	298	94
Target	Clay	High	82	78	95.1	78	95.1
Target	Rubber	High	125	118	94.4	119	95.2
Target	Saline	High	110	98	89.1	101	91.8
Target	Bulk	High	201	185	92	188	93.5
Target	Sheet	High	116	109	94	110	94.8
Pseudo-target	Sheet	High	10	10	**100**	10	**100**
			Num Non-targets	Num FAs	PFA [%]	Num FAs	PFA [%]
			1371	163	**11.9**	15	**1.1**
				Num Scans with FAs	Avg Num FAs	Num Scans with FAs	Avg Num FAs
				110	1.57	15	1

No special rules:
93.6% / 11.9%

12 new rules
95.1% / 1.1%

Additional rules can lead to
100% / 0%

This is similar to how vendors train to pass the test

Lawrence Livermore National Laboratory

42
LLNL-PRES-663046

Thank you!

- We appreciate being involved in this project with other talented researchers
- Special thanks to Carl, David, Clem and ALERT the team for their guidance and patience
- Laura Parker of DHS for her support of our involvement

Lawrence Livermore National Laboratory

43
LLNL-PRES-663046

Lawrence Livermore National Laboratory

11.5.2.8 Discussion, next steps (Crawford)

"Discussion & Next Steps"

Discussion & Next Steps

Carl Crawford, Csuptwo
David Castanon, Boston University
Clem Karl, Boston University
Harry Martz, Lawrence Livermore National Laboratory

1

So What? Who Cares?

- So What? ... What was done ...
  - Five ATRs developed: PD ~ 90%, PFA ~10%
    - Targets: saline, modeling clay, rubber sheets
    - Scanning: Medical CT; single-energy, 500 target scans
  - Automating scoring tools developed
    - Ground-truth labels - semi-automatically
    - Standardized reports
  - All of the above in public domain, by request
- Who cares?... To be determined by you if true ...
  - Problem maps to security scanners
  - ATRs novel with respect to literature in public domain
  - Researchers available to contract to vendors
  - Students trained to work in industry
  - Third parties can work on unclassified, relevant projects
  - Scientific method continues to be applied for more improvements
  - TSA deploys better equipment derived from this project

2

How Good Did They Do?

- Gulp!
- Very difficult question to answer may be because:
  - Medical CT scanner
  - Detect benign objects
  - Vendor ATR are proprietary
  - Literature may not represent vendor's ATR
  - Did not take EDS certification at TSL along with CRT and excursion testing
  - PD/PFA 90%/10% may not be aggressive goal
  - Too few difficult cases
  - Automatic scoring with precision/recall too difficult
  - Comingled PD/PFA runs

3

Common Strengths

- Understood
  - Requirements for ATRs
  - Difficult problem
  - Issues caused by CT artifacts: streaks, blurring
- Completed ATRs
  - 90%/10% PD/PFA on targets
  - Most of pseudo target sheets
- Very smart people

4

Novel Ideas

- Shape-based classifiers
- Pixel classification followed by region growing
- Random forests
- Parallel segmentation
- Graph based splitting
- MAR using in-painting
- Filter banks
- Avoid curse of dimensionality

5

Opportunities for Improvements

- Researchers have done excellent work.
- Domain experts applaud all their efforts
- Next slides discuss opportunities for improvements
  - Should not be considered to be criticism of their work
- We bear some responsibility for weaknesses
  - Corollary of Heisenberg's Uncertainty Principle is that we could not observe without affecting

6

How Far Did They Go?

- Groups were told to mainly concentrate on streak artifact reduction (mainly caused by metal)
- Other artifacts less explicitly addressed
  - Low frequency shading
- Causes
  - Beam hardening, scatter
  - Finite source/detector apertures

7

Areas of Concern

- Some algorithm paths recreated aspects of known methods
- May have over-trained on data
- May not be independent of shape, size, location and orientation
- May not allow range of densities on saline and clay
- Extensibility not tested
- 90%/10% too easy
- Not enough difficult cases
- Imatron scanner's IQ better than security scanners?
  - Not enough photon starvation and artifacts
  - Resolution higher
- Precision/recall values wrong

8

Algorithmic Futures

- More time to work on algorithms
- Combine methods
  – Example: iterative + sinogram processing
- Test on scans of explosives
- Try to pass TSL certification

9

Researchers - Future

- Publish, patent, present
- Seek additional funding from
  – Vendors, DHS, TSA, ALERT
- Access scans of explosives on security scanners
- Revise algorithms
- Address computational expense
- Work on other modalities
  – Personnel scanners (AIT)
  – Line and sparse-view scanners for check point
  – Cargo inspection
- Protect intellectual property

10

Program Management - Future

- Complete final report
- Database and problem statements into public domain
- Facilitate community and networking

11

Scoring Tools

- Provided standardized methods and tools
  - Learned this from LLNL liquid detection project
- Sample ATR allowed
  - Tool verification at early stage
  - Provided ATR developers with structure for input, output and control
- Ground truth not perfect, but allowed automation
- Rules (precision, recall, detection, false alarm)
  - Mostly fair
  - Missing a target by little reduces PD and increases PFA

12

Database Creation

- More specification and documentation
- More and earlier validation
- Trial runs
- More people
- Frustratingly hard to get this entirely, perfectly right. Much time needs to be given to boring things (like record keeping)
- Less object philosophy issues
- Long documentation available

13

Database Availability

- Data available on FTP site
  - CT images, raw data
    - Some dual energy scans
  - Scanner offline reconstruction program, specs
  - Simulation tools
- Scoring tools
  - Source code
  - Spread sheets
- Only need to sign NDA with ALERT
  - www.northeastern.edu/alert/transitioning-technology/alert-datasets/

14

Recommendations to DHS/TSA

- Fund additional research by researchers, national labs and vendors
- Encourage vendors to engage third parties
- Choose more representative unclassified problems
  - AIT, AT2, cargo

15

Recommendations to National Labs

- Execute reconstruction algorithms on scans of threats and stream of commerce data
  - Use DHS image database at LLNL
- Predict improvement on PD/PFA

16

Recommendations to Vendors

- Compare proprietary reconstruction algorithms to researcher algorithms
  - Run algorithms at vendor sites
- Hire researchers, students and their colleagues
- Contribute to specification of more unclassified problems! Share your ideas for what is valuable and what is not.

17

Program Management Changes

- Will be difficult in future to find a smarter, nicer project team
- Choose researchers from more diverse background
- Better communication on project goals
- Push researchers to consume data earlier
- Fund more projects to conduct at ALERT using the structure of task orders

18

Future Task Orders

- AIT
  – Reconstruction, ATR
- EDS/checkpoint/Cargo
  – Sparse view reconstruction
  – Simultaneous reconstruction/segmentation
- Simulation tools
  – Better scanner models
  – Simulated bag/cargo/personnel databases

19

Thank you!

- DHS
  – Vision and funding
- Northeastern University/ALERT
- Researchers
- Vendors
  – Setting up the project
  – Vendor panel today
- Meeting participants

20

The Structure of Scientific Revolutions
Thomas Kuhn

Kuhn has made several notable claims concerning the progress of scientific knowledge: that scientific fields undergo periodic "paradigm shifts" rather than solely progressing in a linear and continuous way; that these paradigm shifts open up new approaches to understanding that scientists would never have considered valid before; and that the notion of scientific truth, at any given moment, cannot be established solely by objective criteria but is defined by a consensus of a scientific community. Competing paradigms are frequently incommensurable; that is, they are competing accounts of reality which cannot be coherently reconciled. Thus, our comprehension of science can never rely on full "objectivity"; we must account for subjective perspectives as well.

Look forward to paradigm shifts in the near future

21

Feedback

- All participants, please fill out questionnaire
  - www.surveymonkey.com/s/TaskOrder4
  - ALERT will email this link
- Other feedback – call/email
  - Carl Crawford
  - Clem Karl
  - Harry Martz

22

So What? Who Cares?

- So What? … What was done …
  - Five ATRs developed: PD ~ 90%, PFA ~10%
    - Targets: saline, modeling clay, rubber sheets
    - Scanning: Medical CT; single-energy, 500 target scans
  - Automating scoring tools developed
    - Ground-truth labels - semi-automatically
    - Standardized reports
  - All of the above in public domain, by request
- Who cares?... To be determined by you if true …
  - Problem maps to security scanners
  - ATRs novel with respect to literature in public domain
  - Researchers available to contract to vendors
  - Students trained to work in industry
  - Third parties can work on unclassified, relevant projects
  - Scientific method continues to be applied for more improvements
  - TSA deploys better equipment derived from this project

23

Good Return on Investment?

- ADSA participants have made direct impacts on vendors' products
  - Many of these people entered the security field because of ADSA
  - Vendors identified performers by participating at ALERT events
- Expect contributions to TSA procurements from this ATR Project, as the previous ADSA and ALERT projects
  - This process will benefit the vendors and the government

24

11.5.3 Questionnaire

11.5.3.1 Questions

"ATR Project – Program Review Questionnaire"

ATR Project – Program Review Questionnaire

Name (optional):

1. What is your relationship to the reconstruction project?
    a. Researcher team
    b. ALERT team member
    c. Current security vendor
    d. Other industry
    e. Government
    f. Academia
    g. Other, please specify
2. What are your comments the approach and results of the ATR research groups?
    a. Purdue University - Charlie Bouman, Dong Hye Ye & Pengchong Jin
    b. University of Tennessee - Jens Gregor
    c. University of Wisconsin - Jun Zhang, Laura Drake, Hongquan Zuo
    d. Massachusetts General Hospital - Synho Do
    e. Lawrence Livermore National Laboratory - Philip Top, Ana Paula Sales, Timo Bremer, Harry Martz
3. Which of the algorithms (or parts thereof) of the research groups seem the most promising for further development?
4. How likely is it that you or your company will use the results of this project?
5. What are your comments on the database (e.g., choice of targets, use of medical scanner, and creation of ground truth labels)?
6. What are your comments on the scoring tools?
7. How should third parties (e.g., researchers from academia) be engaged to develop advanced explosive detection equipment?
8. How should the results of third party development be deployed?
9. Do you feel there is value in the availability of a common set of validated security related data and scanner models in the public domain?
10. What changes should be made for future projects?
11. What topics should be addressed in future projects?
12. What are your additional comments on the ATR project?

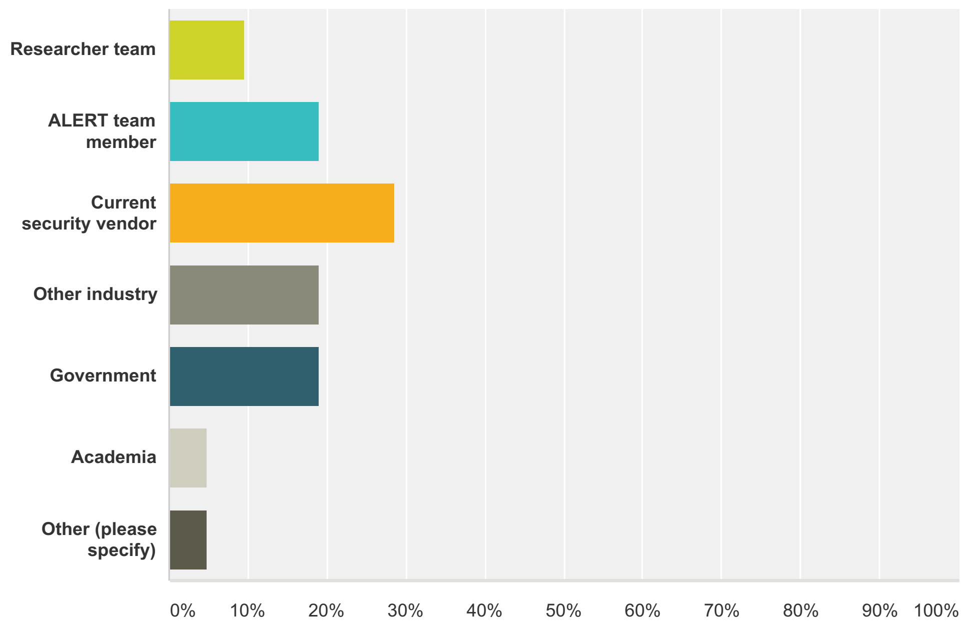
11.5.3.2 Results

"Task Order 3 Program Review – Questionnaire – SurveyMonkey"

Task Order 4 Program Review - Questionnaire SurveyMonkey

Q1 What is your relationship to the reconstruction project?

Answered: 21 Skipped: 0



Answer Choices	Responses	
Researcher team	**9.52%**	2
ALERT team member	**19.05%**	4
Current security vendor	**28.57%**	6
Other industry	**19.05%**	4
Government	**19.05%**	4
Academia	**4.76%**	1
Other (please specify)	**4.76%**	1
Total Respondents: 21		

#	Other (please specify)	Date
1	Independent contractor	11/25/2014 9:53 AM

Task Order 4 Program Review - Questionnaire

SurveyMonkey

Q2 Purdue University - Charlie Bouman, Dong Hye Ye & Pengchong Jin

Answered: 8 Skipped: 13

#	Responses	Date
1	The approach was sound and their results were excellent. Good combination of supervised and unsupervised techniques.	11/24/2014 2:27 PM
2	Good approach with practical results - metal artifact reduction in image space was an interesting approach ... with obvious benefits and risks - appeared to address/limit over-training, but still no objective way to measure - working with clusters and multiple classifiers - larger aperture 450 - 2500 MHU	11/10/2014 4:45 PM
3	Good work.	11/10/2014 12:11 PM
4	Well Done	11/10/2014 11:54 AM
5	Got it and advanced the art	11/10/2014 11:51 AM
6	Good approach. Addressed ATR from segmentation and image correction through discrimination. Had several stages of automated action (separation of bulk and sheet, final discrim, and I believe one or more others) trained on the limited data set. This provides a lot of opportunity for inadvertent over training.	11/10/2014 10:13 AM
7	interesting results on clustering data try using SVM without kernels first to better understand structure of the features and to reduce over-fitting	11/10/2014 10:04 AM
8	Some very good ideas. Took comprehensive approach to the problem	11/7/2014 11:26 AM

Task Order 4 Program Review - Questionnaire

SurveyMonkey

Q3 University of Tennessee - Jens Gregor

Answered: 6 Skipped: 15

#	Responses	Date
1	The approach was sound and the results were good. Did not seem very robust. Too easy to have large shifts in the results.	11/24/2014 2:27 PM
2	Concern over small inclusion zone for voxel density (lower limit 900 MHU)	11/10/2014 4:45 PM
3	Entry level training program.	11/10/2014 12:11 PM
4	Well Done	11/10/2014 11:54 AM
5	Inventive, with a very small team	11/10/2014 11:51 AM
6	Use of graph based object splitting was novel. Recognized the need for "special case" handling rules. Limited discrimination presented. Probably cause of high FA	11/10/2014 10:13 AM

Task Order 4 Program Review - Questionnaire

SurveyMonkey

Q4 University of Wisconsin - Jun Zhang, Laura Drake, Hongquan Zuo

Answered: 6 Skipped: 15

#	Responses	Date
1	The simple modular approach has strengths but merged samples were missed. Need more robust approach.	11/24/2014 2:27 PM
2	Going to look into MRF segmentation and Mean Field Theory Concerned about the computational cost of the approach that was described Concern over small inclusion zone for voxel density (lower limit 950 MHU?)	11/10/2014 4:45 PM
3	Entry level training program.	11/10/2014 12:11 PM
4	Did not see presentation	11/10/2014 11:54 AM
5	Seemed quite traditional	11/10/2014 11:51 AM
6	Use of histogram of all voxels is very general, but subject to problems. First, the histogram approach ignores spatial distribution. Consider several spatially separated objects with overlapping densities along a continuous band of density. This would likely result in an odd combined histogram, requiring a lot of iterative refinement, where a more spatial grouping assessment would immediately see that they are separate articles.	11/10/2014 10:13 AM

Task Order 4 Program Review - Questionnaire SurveyMonkey

Q5 Massachusetts General Hospital - Synho Do

Answered: 6 Skipped: 15

#	Responses	Date
1	Was not there for this presentation	11/24/2014 2:27 PM
2	Concern about over-training	11/10/2014 4:45 PM
3	Not serious.	11/10/2014 12:11 PM
4	Poor effort	11/10/2014 11:54 AM
5	Did not complete the task	11/10/2014 11:51 AM
6	The approach of establishing a density band for each known target, then doing a CCL on all objects that fall in that band is problematic once you get more than a few materials. Computationally, this would take quite a while to run through. While this would prevent a case of an object at one density bleeding onto an object at a substantially different density, in the end, it seems unlikely that it would result in significantly different results than a basic, gray level CCL.	11/10/2014 10:13 AM

Task Order 4 Program Review - Questionnaire SurveyMonkey

Q6 Lawrence Livermore National Laboratory - Philip Top, Ana Paula Sales, Timo Bremer, Harry Martz

Answered: 8 Skipped: 13

#	Responses	Date
1	The effort made to try to increase detection by training on the test set was both surprising and appalling. One might conclude that the team did not differentiate between "acing the test" and designing a generalized system.	11/25/2014 9:11 AM
2	Was not there for this presentation	11/24/2014 2:27 PM
3	Interesting idea about the ensemble approach, analysis of slabs used to create larger objects	11/10/2014 4:45 PM
4	Entry level training program.	11/10/2014 12:11 PM
5	Impressive	11/10/2014 11:54 AM
6	Seemed not to understand their own work	11/10/2014 11:51 AM
7	I would have liked to have more technical description of the segmenters, as they were the primary novel part. The presenter did not seem prepared to discuss.	11/10/2014 10:13 AM
8	A few good ideas, like the parallel segmentation that gets eventually merged. Too bad that the author was not there	11/7/2014 11:26 AM

Task Order 4 Program Review - Questionnaire

SurveyMonkey

Q7 Which of the algorithms (or parts thereof) of the research groups seem the most promising for further development?

Answered: 6 Skipped: 15

#	Responses	Date
1	The idea of multi-classifier is very nice and useful. Training sets or data seem to be very important for ATR. In these images, the noise level is relatively low. The detectability of the target are very much depending on CNR. The noise property is not considered except the implicit work by training process. To be optimal, the reconstruction operation might be combined together in ATR. Also, metal artifacts is an important issue here. Imaging domain impainting might not be enough.	11/27/2014 4:54 AM
2	Purdue was very strong. Would be useful to put the various groups together now and see if an optimal approach could be used. Building on each other.	11/24/2014 2:35 PM
3	Metal artifact reduction (modification of the approach presented by Purdue) Clustered classifiers Slab analysis	11/10/2014 4:57 PM
4	So far they haven't tried serious new ideas that we don't know/practice.	11/10/2014 12:20 PM
5	LLNL was tantalizing, but without enough details to judge. Purdue was consistent with much of what is currently done. Others weren't really practical	11/10/2014 10:22 AM
6	pre-processing and feature construction. Do not fully believe in learning results as they are built on small data sets with methodologies that do not fully prevent over-fitting. for example, 1) pre-processing and segmentation was developed on the full set. 2) as far as I noticed, none of the researchers volunteered to sequester a final test set. 3) some of the pre-processing (clustering) involved shape features that might break the scoring rules	11/10/2014 10:14 AM

Q8 How likely is it that you or your company will use the results of this project?

Answered: 5 Skipped: 16

#	Responses	Date
1	Not likely	11/24/2014 2:35 PM
2	We will definitely try some of these approaches, or derived ideas that were stimulated by these presentations	11/10/2014 4:57 PM
3	Not likely	11/10/2014 12:20 PM
4	Unlikely	11/10/2014 10:22 AM
5	It is possible	11/7/2014 11:29 AM

Task Order 4 Program Review - Questionnaire

SurveyMonkey

Q9 What are your comments on the database (e.g., choice of targets, use of medical scanner, and creation of ground truth labels)?

Answered: 7 Skipped: 14

#	Responses	Date
1	The database was reasonable for this stage of development. Eventually the container will be very important, ie: suitcase hardware.	11/24/2014 2:35 PM
2	Good surrogate problem given the limits of classification and sensitive data. There should have been a blind aspect to this to get more objective measures of algorithm performance.	11/10/2014 4:57 PM
3	Good starting point	11/10/2014 12:20 PM
4	should have used real bags and checkpoint bins should have used broader range of threats should have collected dual energy data nice job on ground truthing, extremely tedious, possibly overkill	11/10/2014 12:00 PM
5	Reasonable database given the constraints of the security restrictions. Sheet targets should have been better handled. Dealing with sheets often drives the need for creative segmentation and discrimination; however, "Sheet" targets are only interesting in that they challenge the resolution of the system. The sheet targets should have been selected to challenge the surrogate system's ability to image them or images should have been degraded to reflect the problem. As collected, with the resolution of the system, the "sheets" were really more of a "thin bulk" which does not require the same level of innovation.	11/10/2014 10:22 AM
6	a good collection. Need larger sets of random unlabeled luggage to prevent over-fitting.	11/10/2014 10:14 AM
7	Given the constraints, the ALERT team did a great job in selecting the targets and using the medical scanner. Some more thought could have been put into the configurations, though (for example the sheet thickness was too high for the medical scanner resolution)	11/7/2014 11:29 AM

Task Order 4 Program Review - Questionnaire

SurveyMonkey

Q10 What are your comments on the scoring tools?

Answered: 7 Skipped: 14

#	Responses	Date
1	Appears to have been well thought-out.	11/24/2014 2:35 PM
2	Good tools and scoring infrastructure. Limits on recall and precision were likely to be too relaxed ... particularly concerned about recall as low as 0.1 - 0.2. Suggest an individual recall limit for each target based on degree of difficulty (e.g. a sheet next to a similar density bulk item in presence of metal streaks might have a lower recall than the same physical sheet target in a clear field).	11/10/2014 4:57 PM
3	Report performance based on expectation in real operation. For example, How ,any bags has false alarm, in addition to how many innocent objects.	11/10/2014 12:20 PM
4	the requirements for precision and/or recall were probably too easy	11/10/2014 12:00 PM
5	Reasonable. For future work, effort should be expended on tightening up the truth mask and requiring a higher overlap with the target	11/10/2014 10:22 AM
6	good	11/10/2014 10:14 AM
7	Very nice work.	11/7/2014 11:29 AM

Task Order 4 Program Review - Questionnaire SurveyMonkey

Q11 How should third parties (e.g., researchers from academia) be engaged to develop advanced explosive detection equipment?

Answered: 6 Skipped: 15

#	Responses	Date
1	There is little incentive for vendors to invest the time and money in improving their systems. Some things to consider might be: seed contract, purchase commitment from the buyers, tiers of certification, and "reward" for better performance.	11/25/2014 9:15 AM
2	Not if they do not have the clearances needed to fully understand the problems and issues.	11/24/2014 2:35 PM
3	Use this and similar forums to expose people and ideas, then let market forces take over. Unfortunately, with limited government funding flowing into this area and reduced procurement budgets to generate business and fund internal development it is less likely for funded partnerships to flourish.	11/10/2014 4:57 PM
4	Become first party.	11/10/2014 12:20 PM
5	collect more extensive data sets that can enable validation of 3rd party algorithms	11/10/2014 10:14 AM
6	Now that the presentation is done, the government should let the industry reach out to the third parties and fund some efforts to put some of the concepts in practice	11/7/2014 11:29 AM

Task Order 4 Program Review - Questionnaire
SurveyMonkey

Q12 How should the results of third party development be deployed?

Answered: 6 Skipped: 15

#	Responses	Date
1	Funding with incentives. There needs to be a government purchasing mechanism which does NOT favor status-quo, least-common denominator, and instead rewards innovation and improvements to designs, performance, sensitivity, resolution, etc. Otherwise, the current model of lowest-cost destroys incentives and delays progress.	11/24/2014 5:07 PM
2	It is good to see if they come up with any insight, but they should be working with the product developers after the initial studies. The good ideas should be transitioned to the manufacturers.	11/24/2014 2:35 PM
3	Deployed in partnership with OEMs as part of EDS, checkpoint or cargo solution.	11/10/2014 4:57 PM
4	This should be considered as an educational effort.	11/10/2014 12:20 PM
5	DHS should issue grants to vendors willing to work with one or more 3rd parties. There is no incentive for the vendors to do this on their own.	11/10/2014 10:22 AM
6	Government should provide requirements to the vendors to enable independent testing of 3rd party algorithms.	11/10/2014 10:14 AM

Task Order 4 Program Review - Questionnaire

SurveyMonkey

Q13 Do you feel there is value in the availability of a common set of validated security related data and scanner models in the public domain?

Answered: 7 Skipped: 14

#	Responses	Date
1	Yes.	11/27/2014 4:54 AM
2	Definitely.	11/24/2014 2:35 PM
3	A common set of data is useful as long as there is a common and independent process for interpreting and measuring results when candidate algorithms are applied to this data.	11/10/2014 4:57 PM
4	It has value and it can cause harm to security.	11/10/2014 12:20 PM
5	NO. The government should not be sponsoring the collection and posting in the public domain of this type of data, depending on how closely related it is to the hardware used to collect the data, the targets, the software, and the results.	11/10/2014 12:00 PM
6	Yes. If nothing else, it increases the qualifications of candidates for hire.	11/10/2014 10:22 AM
7	absolutely yes	11/10/2014 10:14 AM

Task Order 4 Program Review - Questionnaire

SurveyMonkey

Q14 What changes should be made for future projects?

Answered: 4 Skipped: 17

#	Responses	Date
1	Independent scoring and/or comparison of results.	11/10/2014 4:57 PM
2	Tune to areas thye can contribute more. (imaging modeling).	11/10/2014 12:20 PM
3	Get a classified project going so that (cleared) skilled people can solve the real problem working with real data.	11/10/2014 12:00 PM
4	I suggest continuing ATR project as it shows a great promise. So far, the project produced a lot of interesting preliminary results but researchers are only approaching the hard questions - which features will work, which targets are hardest, what will be operational performance. To build on the first step: 1) absolutely keep 2 sets of sequestered data to avoid over-fitting. Allow researchers to get test results on the 1st one, say, not more often than every month, and use the last one for final scoring 2) collect all pre-processing algorithms and feature built by all research teams and enable all teams in the 2nd phase to analyze the features, with the focus on robustness of machine learning results. 3) have an ongoing effort to test all algorithm on operational data	11/10/2014 10:14 AM

Task Order 4 Program Review - Questionnaire SurveyMonkey

Q15 What topics should be addressed in future projects?

Answered: 4 Skipped: 17

#	Responses	Date
1	Go back to segmentation ... it was clear from this ATR workshop that segmentation is critical to detection scores, yet none of the teams worked with earlier segmentation team members.	11/10/2014 4:57 PM
2	Modeling projections and reconstruct CT images.	11/10/2014 12:20 PM
3	ATR in AT2 ATR in AIT	11/10/2014 12:00 PM
4	develop methodology for testing segmentation and ATR on larger and more operatoinal sets	11/10/2014 10:14 AM

Task Order 4 Program Review - Questionnaire SurveyMonkey

Q16 What are your additional comments on the ATR project?

Answered: 3 Skipped: 18

#	Responses	Date
1	Would have improved ability to evaluate results if scores on an independent "blind" data set were provided.	11/10/2014 4:57 PM
2	Good effort.	11/10/2014 12:20 PM
3	Overall I think this was successful, a lot of work, well-executed. But there is still a veil between this project and real problem with real explosives and real bags. And there is less of a problem in detection for checked baggage than there is for checkpoint or cargo, so, while the scope of this project was appropriate, it didn't address the bigger problems at hand.	11/10/2014 12:00 PM

11.5.3.3 Minutes

"ATR Program Review Minutes – November 6th, 2014"

The program review minutes were edited for purposes of clarity. All errors in the minutes are due to the editors of this report and not due to the speakers themselves.

ATR Program Review Minutes – November 6th, 2014[1,2,3]

Speaker: Carl Crawford

Q: The process of the teams to develop the work and algorithms, did they do sensitivity testing on the algorithms?

CC: No. We decided to take the reconstructions off the scanners and use them. We had another project a year ago for reconstruction algorithms. There is a final report for that and we can send the link. It did limit the views coming off the scanner to about 20. We learned that we have to segment the project to look at the scope.

Q: That tack is something that we are formulating right now, including sparse-view reconstruction and sensitivity.

Q: (???)

CC: I would agree with that completely. (???) about detecting benign objects (???).

Q: The bottom line is you have met certain criteria to pass the test.

CC: Right. (???) It doesn't take away.

Q: Talk about the multiple datasets, including the ones that are suitcase oriented.

CC: Right. We didn't use (???).

Q: It would be a mistake to match (???) 90%. The results depend on how you want to build your own best case scenario. Try them on realistic scenarios.

CC: Yes. The numbers are meaningless here.

Q: Glass beads in clay, did you always put them in clay or sometimes?

CC: Sometimes. There is a database/excel spreadsheet if you want to find the target and when it was scanned.

Q: As well as packing and unpacking.

Q: Did you characterize the scanner and resolution points?

CC: Yes. That was part of the previous task order. I will not be reviewing that here but I can give you the numbers at another time.

Q: I usually use the points (???).

CC: I will dig it up. It is approximately a 1 millimeter resolution.

---

[1] Q" indicates a question or comment made by a program review attendee.
[2] Inaudible or missing portions of the minutes will be indicated in parentheses as (???).
[3] "A" indicates a statement made during discussion at the program review by an attendee.

Q: You don't have to critically get that. If you look at that and your targets, it gives a feel for how realistic the scenario was.

Q: The quarter of an inch objects it sounds like (???) as far as I am concerned.

Q: Exactly.

Q: Single or dual energy?

CC: Mostly single.

Q: There is a monograph that has been posted and available on the reconstruction data and instrument. They are the same for the reconstruction and target effort. You can garner a lot of the information we are discussing from that.

Q: Did you outline how to separate for training and testing?

CC: We left that decision to the individual groups. If we were to hold back data, we would have to test it here in Boston; we would have to recreate everyone's environment. We ran this on honesty and did not hold them on that.

Q: What do you mean by (???

CC: I will get to that.

Q: The requirements were what again?

CC: 20% and for quarter of an inch, 10%.

Q: Is there any thought moving forward that you would allow them to use the raw data to reconstruction and use the algorithm?

CC: We said at ADSA01 that we will do ATR and reconstruction separately. Future researchers can now use the algorithm for the ATR.

Q: You allow them to use the raw data but not for the full advanced algorithm. Is that for a region of interest?

CC: Yes.

Q: They still did not take advantage of that?

CC: Yes.

Q: The key product is the leave-behind for future projects.

Q: They were provided images that were segmented or non-segmented?

CC: Non-segmented.

Q: Was there any beam hardening used in the corrections?

CC: Yes, in one, I believe.

Q: They were provided with truth images for all the images?

CC: Yes. There were 188 scans, each had the raw scan and the ground truth.

Q: There is one per bag, but not a bag alarm.

CC: The scanner we had was limited. We were limited to 200 scans. We ran out of time.

Q: Standard format for image format?

CC: Yes. It is called FITS. It allows for data formats. All the image data was converted to that.

Q: Metadata wasn't put in that, just the image?

CC: Yes. Is it a good problem? Vendors?

Speaker: Franco Rupcich

Q: What does SSN mean?

FR: Scan Serial Number.

Q: That's the dimensions of the object?

FR: Yes, not the dimensions of the container.

Q: What kind of gap did you interpolate?

FR: It is very target dependent. It is one of the issues we saw for a certain target. I will get to that in a minute.

Q: Did you consider volume for the targets to determine the number of pixels in the ground truth image?

FR: Not at this stage. I used a separate program that verified the ground truth based on what we expected for mass and volume from the database.

Q: Can you say more about reducing PR?

FR: This is a case with very thin sheets. They are rolled up with gaps in between. There are streaks and shading. When trying to couture this in the program, it would stick to the edges. The edges are only 1-3 pixels thick. It is thrown off by the streaking. The best I could achieve would be to outline around the whole thing. This would be 100% recall but with not very good precision.

Q: But you did bridge the streak artifact?

FR: This is the final ground truth for these. If there were no artifacts, I could get rid of that area in between. There is not much I could do. We could reduce the PR specs down to 1.4.

Q: Carl, the sheets are the only pseudo targets that were included in the results?

Q: That is right.

Q: You have this ground truth created using these canned algorithms. You have some issues with segmentation of certain objects. The solution you are using is to reduce the PR for those. Is there any metric you used to evaluate if others did better on the segmentation?

FR: We will find out. I don't know. I suspect some of them did.

Q: There are cases where recall is better.

Q: But no metric to evaluate that.

Q: No.

Q: The pseudo target is line 16?

FR: Yes.

Q: Is there sufficient documentation that people can pick those pieces of software up and modify them?

FR: There is a specification for all the programs, inputs and outputs. Yes.

Q: On false alarm scoring, any object that was promoted as an alarm, it was overlapping target even if it wasn't overlapping with one of the targets?

FR: It is a labeled generated by ATR but not recall.

Q: Can you say how you handled the object splitting in multiple parts?

FR: There was one case that was clay snaked around. It looks like three separate circles. It was easy. I segmented around and circled those three images.

Q: Did you maintain the (???) where you could outline the three pieces and then count?

FR: Yes. I outlined the whole thing and it created one mass.

Q: Create the same label for those?

FR: Yes.

Q: But they are three separate targets?

FR: Right.

Q: So that would be missed detection and false alarm. But we all have the same tools.

Q: It is what it is. It's a level of knowing what's on the bottom of that.

Q: The scoring and validation becomes a science of itself. There is a small amount of literature. Here we are collecting data and handing it to a third party. The amount of data is much more difficult. We learned that DNDO has done a much better job than we have done. Harry learned that and communicated that to us. We have to go back to DNDO and learn what they did the next time around.

Speaker: Dong Hye Ye

Q: The metal artifact reduction, is that something you came up with independently?

DHY: That's something I came up with myself, yes.

Q: Did you make a definition of a false structure at the early stage?

DHY: (???)

Q: You knew it was (???) because you knew the label?

Q: Let me make that clear. The sheet/bulk identifier is a support machine. Obviously there are soft categories. To train that support vector machine, you (???).

Q: It's a very complex way to (???).

Q: It could be that a simpler thing could have similar performance. The key idea in that piece is the morphological (???).

Q: It looks like the scientific method is working.

Q: Why did you do it in the middle and not in the beginning?

DHY: I wanted to use this (???) to help the (???).

Q: How do you determine if it's merged or non-merged?

A: It's the same way as the (???) classifier. I trained the (???).

Q: How does that translate to unknown objects? If you train on (???).

A: Related to the training issue, I used the (???) integer to describe it, yes.

Q: When you did the cross validation (???) and all your statistics were computed on (???) in the cross evaluation?

Q: The cross evaluation is fine, but if you only have (???) shapes, (???) of them are fine. (???) Is it more about—you're trying to identify bulk objects—is it more about trying to decide between the bulk object and something you don't know about (???)?

DHY: (???)

Q: In your separation of shape (???) and you going to talk about (???). My question is are you going to talk about how that's not using shape in the (???) feature?

Q: They can use shape all the way up to the classifier. (???)

Q: They already have sheet and bulk up there.

Q: That's the rule up front. (???) This was a subjective area and this is a subjective area in the real world.

Q: Did your decision to separate into four separate shape (???)?

Q: You did (???). There's actually 15 clusters unsupervised. When he actually ran the algorithms unspecific on the test cases (???). One way of thinking about it is there is one big classifier that uses one of its features on this unspecified class. (???)

Q: Did you try your (???) reduction early, rather than late and decided the application was better earlier on?

DHY: Yes, (???) I think the (???) wavelength can help. (???)

Q: You have your super meta-classifier as 15 or something. Why 15? Why not 10 or 3?

DHY: I decided the number of cluster that would percolate annually. (???) I tried different numbers like 10, 15, 20. The (???) 15 gives the best reading, so I used that one.

Q: Quantitatively, if you used just one number instead of two or three how does that (???)?

DHY: (???)

Q: So 10% improvement?

DHY: (???) Higher than 5%.

Q: When you did it with a single (???), it was at a rate of 90% to 95%, what you would call a super cluster.

Q: I think Dong tried to do as honest a job as he could to cross validate. But there's always a possibility of some leakage. I don't want to misrepresent it, but I think Dong did as honest of a job as he could to cross validate.

Q: So you went from 4 to 6 to 8 to 12, as you did that as the number of your (???) increased. Did you see a change in your (???)? Did it jump?

DHY: (???) 10, 15, and 20 it didn't change (???).

Q: Did you do anything (???) you tried your morphological position (???)? For example, (???) reduction. If you removed that from the process, did you get an idea of how it changed to get an idea of how your (???)?

DHY: The (???) were lower (???) the segmentation. Our data set was (???).

Q: So you're saying in your data set it wasn't as (???).

DHY: If we have a (???) artifact, then (???) reduction definitely helps. But for a simple, easy case (???) is better.

Q: Was it a binary CCL?

DHY: It was a binary CCL.

Q: CCL took one object and split into two?

DHY: Yes, I found some cases. Even though I set a very wide margin, there was a very limited number of cases. Just a few number of cases. Generally, I didn't see that split in the cases.

Q: That (???) you see there is not the (???). That is not what comes out of it when you have a ton of data.

Q: What about the many cases when the bulk object is attached to the sheet?

DHY: (???) This is not the only feature. For example, there's a number of (???) in the (???).

Q: How do you determine the number of folds (???) you use?

DHY: (???)

Q: But what if your (???) at the bulk, you're going to open the (???) as well.

DHY: (???) If the morphological opening can (???) split the object as well (???).

Q: How do you find the streaks?

DHY: That's on the next slide. We first estimate the variance. We forge the projection using the radon transform.

Q: Do you need to know the system model? Or any radon transform?

DHY: No, I just used the (???). I think that could affect the (???). I just used a very simple way.

Q: How do you find that? The peaks depend on how many bends you have.

DHY: I will come to that in the next slide. The peaks are very important in the segmentation, and the number of material. We find those peaks based on the histogram. We find the major peaks based on the magnitude of the histogram and we assign histogram bins to the major peaks.

Q: You can provide more detail on this later. I don't think this fully addresses the question right now.

Q: Are the labels over a discrete set?

Q: They're sort of constrained to the one.

Q: It's the [part of a formula]. We can talk about it more later offline.

Q: You used the cluster results in place of the general results? Your classifier never sees the full results?

DHY: Yes.

Q: How many of your shape clusters were for (???) objects? You have one there.

DHY: Two.

Q: How many features do you use?

DHY: I use the same features for the sheet and bulk identifiers here [goes back to another slide].

Q: How does it square with the idea that shapes should not be used with the classifier? There was a discussion at the beginning that shapes would not be.

Q: This is legal by our definition. He can have several pathways to a shape.

DHY: For the (???) classifier, I do not use the shape features for my target classifier.

Q: It's not okay to clear targets based on shape.

Q: To the applicability, if a target comes in in one shape, it's going to be applied to one classier. In one classifier, it is likely to be identified, and if it's assigned to another, it could be rejected.

Q: It's a potentially slippery slope.

Q: I can also see how it could be helpful to eliminate false alarms.

Q: (???) You could set the possibility of detection to zero for one classifier based on shape. One thing you might want to do is within each category, try to control the rate of false alarms. We're less constrained because no one is going to be hurt in this situation. We have the opportunity to explore a wide array of ideas.

Q: Let's leave it for later. Let's go through the algorithms and discuss the implications later on.

Q: You fed in the median and the histogram as well?

DHY: Yes.

Q: Okay.

Q: These ten examples that you chose, Carl, are these separate from the training data set?

Q: There was one date set that was given to everybody. These were ten of the 500 targets we picked.

Q: So you used these data sets to build your model?

DHY: (???)

Q: I understand that. Did you already have your model on the end? Are these ten examples you hadn't seen before?

Q: No, they had seen them before.

Q: So this isn't a validation data point?

Q: No, these are all just constraints on the (???).

Q: This is one of those pseudo targets?

DHY: Yes.

Q: Pseudo targets is (???) grams. My point is the target of interest in the ground truth and the target of interest in the (???), you have a very small thing.

DHY: For the (???) target we have a more general.

Q: The threshold is 10%.

Q: Overall, he only has 15%.

Q: Also, remember there are errors in the ground truth. There's some error bars floating on this whole thing.

Q: If you had been able to use the projection data set, these errors could have been avoided"?

DHY: (???)

Q: Can you actually just switch back to your detail scoring sheet while we ask any other questions?

DHY: Sure.

Q: By your metal artifact reduction is doing a better reconstruction, and I think knowing you, that was actually what you were thinking as you went through this. And the filter you're building includes all three elements.

Q: PFA in real life, you don't get one alarm. That false alarm number is not realistic. The way you set up the data, you have to pack your extra alarms in.

Q: Dave is just saying that batting percent is not the percent the vendor comes home to.

Q: How many predictor variables are there per classifier feature that you end up using?

DHY: Between 10 and 30. I cross validate the number of features.

Q: How many features did you end of using?

DHY: It varies depending on the training.

Q: I'm thinking that the impending (???) might not be necessary for your method. All the intensity information is already within the (???) pixels, right? You just use the mask to determine if it's (???).

DHY: (???)

Q: I know that part. When you do the painting, you only get the intensity of the mask area. It's artificial information.

Q: It also prevents him from segmenting more in the later steps.

C: If he wanted to avoid segmentation, he already has mask information.

Speaker: Jens Gregor

Q: So you got a pretty high number on the sheet compared to your predecessor overall and a higher number on PFA. Did you focus on the sheets?

JG: Probably, self-consciously. (???) Many of the difficult sheets were the ones with bulk objects sitting on them. To a large extent, the saline that gets mixed in with the pure water (???).

Q: The low concentration saline is similar to water? Does anyone know the measurements?

Q: 1000 and 35.

Q: You might want to contrast yourself explicitly from Dong He.

JG: I was trying to do that. He got really impressive numbers. Mine is much simpler. I'm working with the raw data.

Q: Both are acceptable.

JG: Right, right. It's just different.

Q: Would it matter if you did density first then shapes?

JG: I've done that, but the density needs so much more data upfront.

Q: Are neighbors adjacent?

JG: Yes.

Q: What do you do for the pseudo target sheets?

JG: They're not split off. They come out the way they come out, if the bulk is not attached to it.

Q: So in your cartoon, if there's equidistant for the sheet, but not for the bulk, how does it split off?

JG: The sheet growth will try to grow its way out this way, while also growing the bulk out. They're growing toward each other. That's the abstract idea.

Q: Does the sheet detector scale well down into the 1, 2, 3 pixel range?

JG: I don't recall having any pseudo detector sheets (???).

Q: Are you using Mass as one of the features?

JG: Yes, mass density.

Q: Mass can be whatever you want it to be. I thought it was disqualified.

JG: You can do minimum on that.

Q: At some point you are not worried about it.

Q: What is the training for those who are doing this work?

JG: There are about 500 objects that are volumetric images that are ground truth images.

Q: Does that on the left bother you?

JG: You have different parameters you can control, and you can control that with the variance, and you can set it up to go in and carve it out. You have to classify it, and it wouldn't look like this.

Q: How do you recall this decision during training?

JG: I have to do it before training. I use the ground truth to find out which ones are close enough to ground truth.

Q: How many bags?

JG: 180 bags. I can do 20 seconds per bag. Some are really simple, and some have a lot of data.

Q: What made you choose SVM over other approaches?

JG: When you look at the data, it's clear that you are not going to be given a separation. The alternative is fine, but I can't see what's happening and why. I can change the parameters with this one.

Q: Other than file IO, is the work memory band restricted?

JG: The only thing that takes a little bit of time is when I find a sheet. I have to find two vectors that are 180 degrees against each other. They don't have to be accurate. I guess there is one thing that I forgot to tell you. When I have something with a lot of pixels, it's very costly. If I had to sit and wait for it, it would take too long.

Q: Would the next step be trying an ensemble approach? Maybe different algorithms together?

JG: There's no money for that. This is a finite amount of money for this.

Q: It seems like you are missing one powerful tool that you could use, which is the image grading. Have you looked into that?

JG: I would be worried about the effect, but I don't know. Maybe once it's segmented out.

Speaker: Jun Zhang

Q: You are doing the clustering before, and a lot of people are doing it after?

JZ: Yes. It's the main part of the segmentation process.

Q: How do you determine how many clusters?

JZ: We will discuss that later.

Q: Is it correct that you are essentially looking at the inverse of what other people are looking at?

JZ: I haven't really understood all of the aspects that we are discussing. We say whether or not it is a false alarm. Sometimes a false alarm also survives.

Q: Does this reduce to kamian clustering?

JZ: It does.

Q: How many classes?

JZ: I will explain that later.

Q: Are you applying this against the entire bag image?

JZ: Good question. In theory, what we are doing is producing an initial estimate; sub-sampling the whole bag. Then the actual EM algorithm will scan everything at full resolution.

Q: Can you see more on the gradient?

JZ: At each pixel, we sum it up and divide by the total number of pixels.

Q: You have no aggregation, so if something is smaller than the mass, you don't care about that?

JZ: Yes. In general we have so many features, we should shrink them to see which ones are really needed.

Q: When you look at the whole shape, it's hard to differentiate different classes, but when you look at the peaks you can see clear differentiation.

JZ: Yes, that's a good point.

Q: What is your average run time?

JZ: On the run time, generally, EM algorithm has 2 versions. One is on the EM itself, the other is on the theory, and each on average 20. 100 passes over the whole volume. It takes several hours.

Q: What is the thought about doing this classification beforehand? You mentioned pushing the probability beforehand.

JZ: You can also use that to determine whether the region is viable or not. Maybe it should be split. If I understand the question, it was about what he presented as far as the PD and PFA at the presentation stage. I think that's critically important.

Q: At least in our current approach it's not there. The next step is discrimination. You want to have the maximum set coming through.

Q: Everybody segments once. If you lose it, it's not just in his approach, it's in every approach.

Q: Yes, part of it is very similar to what is being discussed here.

Speaker: Synho Do

Q: Where is (???)?

SD: It is not here. I think I missed it. Let me try to include it in my final report.

Q: This is the ground truth here?

SD: Yes. This is all ground truth.

Q: You are saying target peak HU values?

SD: Yes, I am.

Q: When you say you increase the number of targets, what do you mean by that?

SD: I had 93 targets in my datasets. In terms of materials, I may have only 4 or 5 materials but I like to detect all of the targets.

Q: Are you making non-targets (???) or false alarm targets, or using materials in different ways?

SD: I rate my target based on the frequency that we put it inside the bins.

Q: There were 500 or so targets and 185 cases.

SD: If we add these then that is the 500.

Q: So these are the items showing up again and again and used in different cases?

SD: Okay. I say target here. That is the item I need to detect.

Q: So 93 things but when put in the bags, the total number of times they were run was 500?

SD: Yes.

Q: So if you add more targets, the false alarm rates go down?

SD: Yes.

Q: He is matching so many histograms. The more cases he uses for potential candidates, the false alarm rate goes down.

Q: Did you detect it or not in this slide?

SD: I think I missed it. In this slice, whether I could detect it or not, I am displaying it. I picked one of the cases.

Q: If you are just using histograms, how do you segment out objects?

SD: By ground truthing.

Q: By using the ground truth image?

SD: I need to detect accurate histogram peak. To do that, I use the ground truth mask to capture all the three structures out of the cases and using the histogram to pick the H value and band width.

Q: Can you walk through the steps from having a bag to what you do to the set of boxes to decide and alarm.

SD: Before I scan, I should do the pre-scan. I should do a few scans in different locations.

Q: New bag.

SD: I am going to detect saline. I pick one from my table the saline, band width and threshold. Then I have the area decision smaller than 1000 to connect or throw it away. (???)

Q: So if you have a bag and don't know what's in it, you do saline and do the band threshold to check for more than 1000 and if it is, you alarm. (???)

Q: This is basically just segmentation. There is classifier on the back end of this?

SD: We already have the (???) table.

Q: So if you take this, you have the result from the segmentation. This goes to the point of earlier that to grade this, you have to see what the PDP is on the segmentation. That's what he's done.

Q: If your histograms overlap over objects, you might have multiple alarms (???).

SD: My histograms overlap but if the peak doesn't overlap and the band doesn't overlap, then I can capture independently at the core.

Q: But if it does overlap, then you have the possibility of a dual alarm.

SD: This can easily be extended to the histogram.

Q: What do you think the detection on the pseudo targets is?

SD: I don't know.

Q: What do you think?

Q: (???)

SD: I think we can separate those two.

Q: Analogue (???) discrete mixes?

SD: If I know the point of spread function of this system, I can do the composition work. The boundary might be (???) the components could get better.

Q: You have a streak on the left but on the right it is merged, is that because the slices have merged?

SD: It is in 3D.

Q: If it wasn't in 3D, would it be split?

SD: It might be split. (???) In my current structure, I cannot generate this metric which is why I didn't use it.

Speaker: Steve Azevedo

Q: Your goal here is determine voxel connectivity and (???) to a target or to (???)?

SA: It is labeling the image with all possible targets.

Q: Saline (???)

SA: Everything.

Q: The way it is worded it sounded similar to Synho's work.

Q: They are using target as object here.

SA: It is all the objects.

Q: You get a lot of stuff, (???) and we are figuring out how to merge the left and right sides. We didn't know if it was going to ready. We figured out in the 11th hour how to merge it.

Q: At this point, it looks like you have targets?

Q: He put a lower threshold and threw a lot of voxels away. We are not going to do any fancy segmentation.

Q: I can't parse that probability part.

SA: There was training data.

Q: Are you using it to compute the probabilities?

Q: You are taking a 10x10 slab. You are deciding whether it is each one of those or any one of those?

SA: Any one of those.

Q: So you are saying it is saline or clay?

SA: That comes later in the classification.

Q: Are you saying it falls between 900 and 20000 and it could be one of these? It seems as though here you are implying that it is part of clay or saline and a conscious decision.

SA: It is general (???) target.

Q: (???) in that slice.

SA: In that slice, in any dimension.

Q: Are you sampling the entire of volume in this way?

SA: It does some parsing out in the beginning. But virtually, yes.

Q: How do you step through the volume when you have three different planes?

SA: (???) Yes, the whole volume.

Q: When you make a comment about narrow channels to different regions, does that illuminate sheets? Sheets themselves are narrow channels. I would think if you are building in the existence of a narrow channel that will make something separate, that seems to make (???) disappear.

SA: That is why there is a separate channel here for sheets. These are very thin objects.

Q: How close were the (???)?

SA: That is the cross validation.

Q: You have to say how much they vary.

SA: That is something you have to ask (???). I don’t know the answer to that.

Q: Steve, did you discover that or did you (???)?

SA: You could.

Q: We don’t know what the targets are until we get there but will your algorithm (???) objects that aren’t in the field of interest?

Q: We discussed that. (???)

Q: We have a lot of common types so we want to classify those things, some very well-known objects first.

Q: The idea is you have a big streak that splits the water bottle. You can compare those that do and don’t. You can see what is consistent and that it might be a water bottle. We are talking about predicting things. We know that government will say that here are targets, explore what they look like and if they will explode or not. They want us to give them an idea and if it will cause an alarm or what if it is surrounded by all these known items.

Q: For each of those 10x10 slabs you are calculating features.

SA: I sent an email to Philip. We had a thing called a pre-classifier. Phil does reclassification and decides if they will be targets or not. You have one set of labeled segments and go through a classification there. That classifier has similar features.

Q: Is segment part of the preclassifier?

SA: These are not slabs. These are the 10x10.

Q: It says slabs there.

Q: If I have an aggregate object, do I have classifier means? We are classifying candidate objects for alarms and false alarms. How many means do I feed into the classifier for that object?

Q: You are getting rid of the artifacts. You are getting rid of the confusions.

Q: Not completely.

Q: Suppose they compute the objects in the slabs, then there is the common deviation (???)

Q: How are you (???) these slabs?

SA: It is an ensemble average of the features within the segment.

Q: If you have three different objects, how do you connect this to (???)?

Q: He has broken them into different objects and treats them differently.

SA: Each 3D (???) is like its own object (???).

Q: You must make a decision that the slab is a part of the segment.

SA: Yes.

Q: Can a slab be part of multiple segments?

SA: (???)

Q: What is the object at the end of (???)?

SA: This is all 3D.

Q: This picture shows that (???)

SA: It shows just (???) slice.

Q: You are taking the probability that it could be one of these targets over slabs and you are joining that with voxel based (???). You have metrics with these slabs that you take with you and get rid of objects, join objects and (???) through the ensemble (???)

SA: (???)

Q: What were the dots in the previous slide?

SA: The dots are highlighting the target.

Q: Green in the ground truth that is not the target.

Q: The maroon is of interest.

Q: But the green?

SA: Other target objects.

Q: How do you make the over trained judgment?

SA: Philip looked at why he missed some targets.

Q: That is overtraining.

SA: Harry did a tough thing to researchers. He said forget what Carl says. See if you can 100/0. Vendors have to get to that point. See if you can and what you have to do to get it.

Q: Vendors go through several steps to pass the test. You learn things in that process. When we learn something we try to take advantage of it. They didn't want to do it. I told them they need to do this. It is forcing you to use not just physics.

Q: But you want to get a perfect score in training data. You can but it takes a lot of practice.

Q: I disagree with your characterization of how the test happens. You mention the key three stages. There is temptation to squeeze for 100/0. You can do that but when you get to the next stage you will fall on your face because you have over trained and it won't deal with new stuff.

Q: Then what do you do from CRT to Pre-Cert?

SA: You don't do anything. You have to start over and do it over again.

Q: But it is the same data.

Q: You don't get the data back. You are told you didn't do well on the subset. It is minimal.

Q: We go back to the target.

Q: You can't over train. The test is blind and you don't know what the targets are on the test itself. You just know the data you've collected. This effort is not duplicating what we do because you missed the aspect to the test which is the generalization part. If the test set has configurations you haven't look at yet, you are dead. Then what do you do? You have to go and collect more data.

Q: Harry's point is not completely off.

Q: You collect additional data and you have them adjust a few things here and there.

Q: But the point is you're trying to generalize that data with an unknown target.

Q: Let's now collect three hundred bags, which are now my new test bed.

Q: But you're not going to get to 101.

Q: The first thing Carl said is this does not mimic the real world. Number 2, we were given a problem, you get some data, you don't (???) on it, you go into it and you don't pass, from what I've heard you get a chance to go back in and do it again. Somehow, assuming you get some pieces of information back, you go back to the lab, do some experiments, tweak some things.

Q: The biggest red flag is a hundred zeros. I predict that PFA exists. I go back to the (???).

Q: When you're talking about training, is it more relevant to the positive features or to the false alarms?

Q: In the beginning, if you don't have it into your classifier, you're never going to detect that object. If you go in with 90% of that, you will get 905, at best 85%. You have to go in and do something to make it better. My colleagues kept getting mad at me because I told them to do what it takes. Imagine that this depends on your job.

Q: You report a false alarm about 10% to define the total number of false alarms. What is the scan of bags, at least the number of bags with false alarms? What's the total number of bags?

Q: How many bins were done?

SA: Approximately 188.

Q: (???) Most of the physics features are the same. I said okay, I'll look at the features. Most of the features I can rely on. (???) You pick some other feature. If I use this feature, it will do a better job at getting rid of this false alarm.

Q: You have control over it. You can always get a hundred and zero. Because you can always make ridiculous rules to tweak your findings.

Q: For the slides with the over training, if you actually took those two algorithms forward and applied them to a new set of data over the six months, which would perform better?

SA: The one to the left for sure.

Q: I think it's a good thing that you did that. They give you ten examples and they ask you to discriminate and create a classifier from ten examples. There are circumstances where you can't do a good training, because you don't have good data. We don't know where that boundary lies. And maybe it's something we need to address. ??? It's a pretty important demonstration.

Q: How would you deal with people in this room whose data sets may be very limited?

SA: I make comments like the one I just made.

Q: Because you can never have too much data.

Q: In the past, the vendors collected the data on their own. If you know what data is needed for (???) you can design the data to match this. There are other situations where the vendors are not in control of data collection and it's imposed on them. (???) The test is going to be different than what you're going to get. The data sample is going to be smaller. (???)

Q: Can you simulate your way out of this?

Q: I think this would be a great topic for this group to address. Monte Carlo (???) I think it would be a really good way to focus, cheaply enhance our data sets.

Q: I'm going to make a comment, because I funded this. Carl has raised the approach and I don't expect people to answer this yet. Although we all agree. No one is going to go off and start (???). My world is full of people that don't understand (???). I need ammunition to get us to move forward. That's the world I'm in. Not the technical reality that it takes to get a system up.

SA: Is this a good use of taxpayers' dollars? How do we help Laura with this question?

Q: That also goes for ALERT as well. We have to know how to the (???) even if an algorithm hasn't made it into a certified machine. Getting (???) from the vendors that this work was valuable and there's a use for it; that kind of word of mouth put down in emails to Laura and I (Michael Silevitch) will go a long way in helping us with this.

Q: According to the newer classification document, any images collected on a system by TSA anywhere belong to them.

SA: (???) If I have a system that I am selling to a government that has other requirements (???) It's very much the same as a medical system.

Q: Can you provide him enough information to train him?

Q: The challenge there is there is a lot of detailed training that happens off the images. Now you need to detect a chunk of some unknown explosives and his system has to go through a training phase on that. You can send him parameters.

Q: How do you turn the knobs?

Q: What specific data? If you send image data vs corrective projection data. There is a segmentation problem. Dong corrected a lot for the image so he could do better segmentation. That is not the case for everybody. This has taught me that the artifacts have a much more dramatic effect in the process. If you start with high quality image data, everything else falls in line. The quality of the data is important.

Q: The challenge with artifact in this field is they don't have time.

Q: (???) about the targets?

SA: We only care about the targets.

Q: What is the real number?

Q: The real number is very similar to what you have. It is not split.

A: I kind of liked when we focused a lot on segmentation up front. I think that they can be structured

A: As far as the 5 presentations, the Purdue one sticks out for me. It will take care of all of the way from segmentation through classification. It sucks to image correction. I would have to see more detail of LLNL.

A: I was impressed with the tools. I was trying to come up with a number also, and the best is a factor of 1000 (more difficult). Segmentation was important, and I wanted all the teams to realize how important that is. The important thing is to detect, and to display. I agree with the ensemble segmentation.

A: I thought it started out the same. It's a hard problem but the results are impressive. I would say that there are pieces of each one of the talks that I found interesting. I won't say what goes in one.

A: This was a hard problem with a hard dataset. I would like to see more than 5 working on the problem. I think you can get good solutions by having a lot of people working on it.

Q: It's something you should think through.

Q: What was wrong? Or what can we do better?

A: Increase the number of bags and bins. Think more about categories as well (expand). We have benign materials with similar characteristics as res.

Q: What is the goal?

A: To approach the complexity that we are dealing with.

Q: Did you think they did that?

A: Yes. They are getting there.

A: The thing that concerned me was that when the separations were in the industry, but in terms of paneling them in the algorithm, there is nothing inherent about those numbers. When you are on a system like a millimeter, it's not as interesting. When one has to deal with the challenges, it has to be relative to what the machine is doing. One would need to think that they are collecting it on this machine. That would be what would makes it better.

A: You knew you were collecting it on a 1 millimeter system. You wouldn't collect sheets on that machine.

Q: The sheets are in the dataset. We can't talk about them, but they are there. They get these high PDs. How do we run a problem like this?

A: There wasn't a lot of point to having ¾ inch sheet in this dataset. You don't have to tie back what a machine can do. Then your score on the sheet would be indicative.

Q: As you replicate the data, do you grow the resolution and see what the effect is? One could represent those things.

A: Even though machines that are going to down that, you see that you are giving one that is more available to you. Are you coming up with anything different.

Q: It would have been nice to have this discussion at the beginning. We have gone through three things. And you know what we can do. What's next? That's how we serve Laura, that's how we serve ALERT.

A: I think I can tie two of them together. Looking back, I began to realize that there was no connectivity. When I look back, I think segmentation is where a lot of good things can come out of academia. The second thing was generalizing. That would have been more unbiased (blind scoring).

A: We didn't have a tool.

A: One that doesn't use tools for ground truth is regulators. That will be vitally important.

Q: How do you emulate that in this kind of test?

A: Pressure.

Q: It's hard to go off and put screw on if you're going to out of money.

A: I don't think you need to put screw on it. (???) I don't know that you want to put pressure on them. I think you want to listen to him and let us take those ideas.

A: Incentivize and come up with a joint proposals. Decide which ones they (???) are working for and which ones are not.

A: You guys can't provide scans to this group for public use, if the different groups can do demonstrations. If we're interested in one or two of them, can you then go and get a BAA and (???)

A: We might not be able to give a full set of images to them. They might have to come to us and we'll have to go to a dark room. It's very difficult from a vendor side to say that's a neat idea, we'll pay. There's not a lot of incentive. I was with a company that improved unilaterally that it performed its performance. I was told by a certain company that it was not going to be able to sell (???) Some of our colleagues have passed it, but it didn't affect (???) It's still; not in the field. When you're doing an algorithm roll, it's not (???) it's millions of dollars to go through that process. You could have a fabulous idea don't go through management (???) It's very difficult. The customers don't go this guy got a 95 and this guy got 93 on the false negative. You should keep in mind on how to move this into the vendor community. The corporation's job is to make money for its shareholders. There's very little incentive to (???)

A: You have to get improved software that taps into the world of what's need, but they have to get incentive.

A: That's not as interesting of a story – but the whole penents of the pint, but for this group we are trying to push for that.

Q: Will you go one step further. Is it true that you would actually like the bar to be higher than the government has made it?

A: A future looking bar.

A: We work for vendors. We are not the company. They perform a service. With a new bar and new investment, which will add more sales, then yes, but if not, then no.

A: I want to clarify that there's no incentive. We are not delinquent. We have all achieved that performance. We have achieved the bar that the government gave us. There is ability to achieve greater than that. Some have a philosophy that we want to be government leaders. We want to go in diff areas and try to push the tech where standards are higher. In terms of performance of machine, there is no economic benefit for it.

Q: The government doesn't want to say what the PD is. But there is more emphasis to look at cost of system and ownership. How many people and through-put. Take that into calculation, what is the cost, then if I lower PD, I get better PFA.

Q: Likelihood that comparison will use ART?

A: Can't answer

A: It seems like you are looking for the quickest way to cert a machine, then you take a cert machine to a vendor that is a risk. I think algorithms should be evaluated. It doesn't have to be with the source code. It's useful in terms of transferability. In terms of things to do in the future, modeling, give us stuff fast to more accurately model bags. That is interesting in terms of minimizing sampling. I think those are my thoughts on the questions. Segmentation step is really detection step. I think that's critical point that we missed. It should be addressed as much. It clears those threats.

The prospects of what you think of cross plant form - an algorithm that could work across multiple machines. I think it would be interesting to know what people think about that.

Robust is critical for us. If you take the same bag and rotate. That if you change parameters of segmentation. Then you can do this problem. If you do that how does it change overall.

A: I want to answer the outstanding questions. You have a lack of a test center. What do we do next? Get Livermore to develop monte carlo model that matches those bags. It would solve the problem of having data libraries. Then you asked about pressure. There are only two ways: flunk or money. Laura talked about *trans* algorithms, you have to show me something that I can't do. You want to make it easier to pass the test.

Q: What about testing. If we collected another set. How do we test them

A: Once you know you have something that's usable, then what do you do?

Q: Game theory. If you do it at the end of the program. Should we have done it blind?

A: Yes.

A: But you don't do it that way.

A: All I am saying is that if you are getting 95/8 and you get 60/30 then you aren't ready to talk.

A: If we do ours and get a score like that, someone gets fired. We don't tweak. You start over. You have to have projected and actual score. As they matured you get your own line set inside. It speaks for each one of these. If all of them hold as they go through. It's likely that some of them would held and some wouldn't

Q: After you do the blind test no. But with those numbers.

A: You structured a project arrangement, this is just assessment. Is it legit? Or is it not?

A: The change of transfer of the effort, is greater than from 90/10 to 60/30. IF we had done this and gotten those results, everyone would be listening.

Q: It seems like the discussion is how we can make the academic look like their own company. Is it the right approach. Where they excel is novel problems. This is maybe addressing that. Even scaled down, this was hard. Is there a way to use resources more effectively.

A: I think it would be neat to have a virtual cert process. If you are doing good research. It should be reproducible. It would rep an honest I did my best to give an honest assessment.

A: All this does is solidify the claim.

A: One thing is that we are not looking for a new segmentation solution. We are evaluating these in our mind. We don't want to change.

A: Even a diff company has a baseline. We are working with TSS on algorithms. So it comes back to that it helps better solidify the claims.

Q: How about the consistency of the dataset? 5% exact? Same vendor/age/etc.? They don't tell you where to get the things to test. How do you come close to the real world, it might need to be broad.

A: It was exact (7%), but they were majority the same, but some were different. We have log files.

Q: Did you scan them by themselves?

A: Mostly yes.

A: They aren't homogenous.

A: Be specific. We had specific rubber. There is nothing that prevented them from saying they can (???)

A: When I was listening to the thoughts I looked at the novelty. I looked at operational capability, and I thought, 'can it be implemented' or does it require revamping. Operational feasibility, with segmentation algorithms, there were no time constraints. We have to buy in a few seconds.

A: It was for 90 bags.

A: That's fine. As long as it doesn't take 100

A: There are CT scanners that don't perform well. You can synth in any direction, and take 2-4 projects, and emulate CT scanners, and see if that improves the check point. That might help the TSA. The check world is better than check point. You might want to look in that direction.

Q: Good use of dollars?

A: Yes. Because you put together a problem with low funds with a set of data that was experimental in nature. They had 5 unique approaches, and people learned something. Will it end up with software in a vendor machine, not sure.

A: It increasing the available folks that had some idea the hiring challenge is that you are used to having these datasets (???)

Q: From 1997, from then, would it be beneficial if we did the back in 1996?

A: Anyone would have been better, that it gave exposure to things that wouldn't work. At the time we were partnered with GE as well. There is no question that it helps from development from scratch. It would have helped TSA. It would have helped with Surescan. It starts the step. Some are better than others.

A: I have been here for a long time and I wish some of this came around years ago. But we do take back that we have learned about from algorithms, I think it's a good use of dollars.

A: Yes, I think it's a good use. As we learned about this, we did not think of all those approaches, not only from vendors but government. In terms of overtraining, the vast approach is that we would try to do it with distributions. There are a variety of things that came out of this, from my point of view.

Q: Did it hurt?

A: It makes vendors look smart. It's a good use of money. It's a step in right direction.

Comments

A: I want to thank the organizing team, Carl, Clem, Harry. It was a monumental effort.

A: Thank you for attending. I think we have found that over the last three task orders, I have taken a lot of notes, and Carl and I and ALERT will discuss next steps.

A: In retrospect, we could have done a lot better.

A: We had limited time and money.

A: Thank you for working on it.

A: And the amount of effort and people that went into this is impressive.

11.6 Researcher Final Reports

11.6.1 ATR Development – Ye

"TO#4: Automatic Target Recognition (ATR) for CT Luggage Screening"

TO#4: Automatic Target Recognition (ATR) for CT Luggage Screening

Dong Hye Ye, Pengchong Jin, Charles Bouman
Purdue University
1/23/2015

This material is based upon work supported by the U.S. Department of Homeland Security, Science and Technology Directorate. The views and conclusions contained in this document are those of the authors and should not be interpreted as necessarily representing the official polices, either expressed or implied, of the U.S Department of Homeland Security.

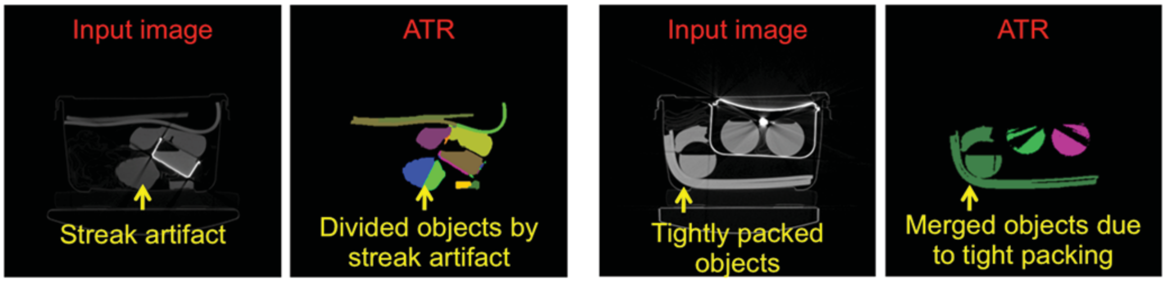
Abstract— In transportation security applications, CT scanners are used to scan checked baggage for threatening materials. The CT scanner will generate the cross-sectional images of the bag. Given CT scans, automatic image analysis for recognizing objects of interest is required by TSA to help extracting important information, and supporting human judgments. A typical Automatic Target Recognition (ATR) system for security applications mainly consists of three steps: object segmentation, feature extraction, and target classification. Our goal is to advance each of these three steps in ATR system. To achieve this goal, we will investigate dataset in deep to understand the challenges in the detection process, design new algorithms for each steps in ATR separately, and integrate them to build up the whole ATR system. In particular, we will study the use of advanced segmentation, feature extraction and classification algorithms to tackle challenging cases, such as images with cluttered objects, artifacts, and in-accuracy CT numbers. During the project period, our ATR systeme achieved 95% of probability of detection (PD) and 8% probability of false alarm (PFA) by meeting the project goals of 90% PD and 10% PFA.

I. PARTICIPANTS

Faculty/Staff (including visiting faculty)					
Name	**Title**	**Institution**	**Email**	**Phone**	**Supported by ALERT Funds (Y/N)**
Charles A. Bouman	Professor	Purdue University	bouman@purdue.edu	765-494-0340	Y
Dong Hye Ye	Post-doc	Purdue University	yed@purdue.edu	215-880-7843	Y
Graduate, Undergraduate and REU Students					
Name	**Degree Pursued**	**Institution**	**Email**	**Month/Year of Graduation**	**Supported by ALERT Funds (Y/N)**
Pengchong Jin	PhD	Purdue University	jin36@purdue.edu	2015	Y

II. PROJECT OVERVIEW AND SIGNIFICANCE

X-ray CT for checked baggage scanning is among the key elements of transportation security [1,2,3,4]. Automatic target recognition (ATR) from the scanned images is required by TSA not to allow humans to perform the first review of images and support human judgments for the second review. However, developing ATR system is challenged due to CT resolution and CT artifacts [5,6,7]. When components of a reconstructed slice of a bag are poorly resolved or corrupted by artifacts resulting from highly attenuating materials such as metal objects, as a part of ATR, poor segmentation of materials may result in ambiguity in the bag's content to require human intervention due to a "false alarm".



(a) (b)

Figure 1. Challenges in ATR for CT images: For (a) and (b), left and right images are input CT images and corresponding ATR masks, respectively. Each detected target mask is individually color-coded. (a) Metal streaking artifacts divide objects detected by ATR; (b) ATR merges objects due to tight packing.

Figure 1 illustrates these challenges in ATR for CT images. In Figure 1 (a), metal introduces streaking artifacts with which ATR detects divided objects. In addition, ATR may not be able to separate cluttered objects by tight packing due to limited CT resolution as shown in Figure 1 (b).

The objective of this research is to investigate and develop a new Automatic Target Recognition (ATR) system that can handle above-mentioned challenging cases. To do this, we will incorporate advanced computer vision algorithms upon the baseline software provided by ALERT. While computer vision algorithms are potential to improve each ATR components, such as image denoising, image segmentation, and object detection, most of them are for the application of natural images. So the questions remain as to the potential advantages of the advanced computer vision techniques in ATR applications.

We successfully developed a new ATR system that incorporates advanced computer vision algorithms such as shape filter and multi-label segmentation. We evaluate the performance using the specified metrics (i.e. probability of detection (PD) and probability of false alarm (PFA)), and it successfully meets the project goal of 90% PD and 10% PFA. With these results, we can propose potential directions for improvement of ATR in aviation security society.

III. RESEARCH ACTIVITY

A. ATR Approach

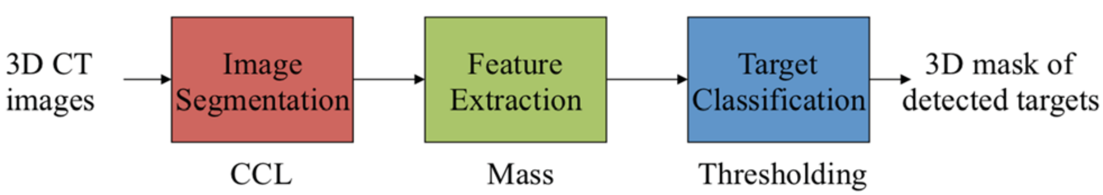


Figure 2. Overview of Sample ATR [8]: Sample ATR consists of three components: image segmentation, feature extraction, and target classification. Sample ATR first segments objects by connected component labeling (CCL). It then measures the CT mass of each segmented objects and detect targets that are heavier than certain threshold.

Typically, an Automatic Target Recognition (ATR) system will consist of several separate processing units, including image segmentation, feature extraction, and target classification (see Figure 2). Sample ATR uses connected component labeling (CCL) to segment objects in the CT scans. It then extracts the mass of each connected component and keeps only object whose mass is higher than target definition. However, sample ATR is limited to tackle the challenges such as streak artifacts and tight packing.

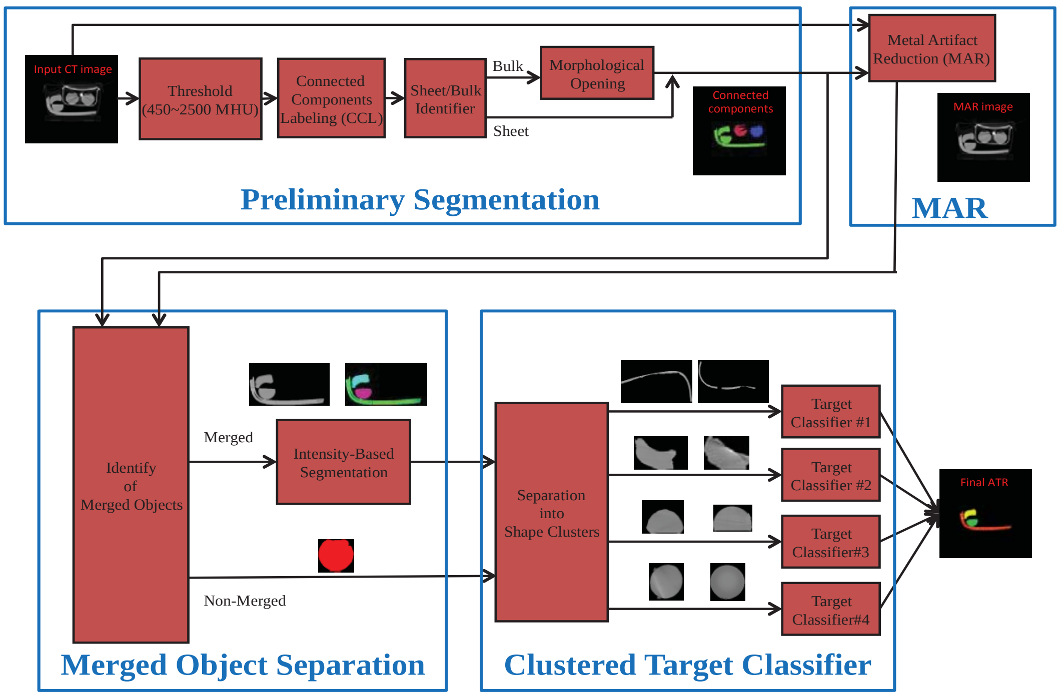


Figure 3. Overview of proposed ATR: First, we pre-segment the image with CCL followed by morphological opening on bulk structure. We then reduce the metal artifacts and further segment merged objects. Finally, we classify each segmented objects with clustered target classifier.

To tackle challenges, we propose ATR as summarized in Figure 3. Given CT scan, we first extract foreground mask through thresholding between 450 and 2500MHU. This range is chosen to cover the intensity levels of our targets even with metal artifacts. Then, we apply CCL on the foreground mask to find connected components. For each connected component, we determine whether the component is sheet or bulk and then apply morphological opening only on bulk structures. This is because morphological opening may erase thin sheet structure. Even after morphological opening, there would be merged objects due to tight packing. So, we perform intensity-based segmentation on each connected components after metal artifact reduction. Then, intensity-based segmentation can lead to over-segmentation due to CT noise. Therefore, we identify merged object based on shape and perform intensity-based segmentation only for merged objects. For each segmented objects, we determine whether each object is target or not using histogram-based features. Then, the features can be overlapped between targets and non-targets which makes difficult to train the classifier. So, we apply clustering the training data set and train the classifier with similar shapes. After target classification, we will have the mask for our detected targets. In following, we will give detailed descriptions about each step of ATR system.

1. Image Segmentation

Selective Morphological Opening on Bulk Objects

Image segmentation is a step in ATR to assess material and morphological properties of the objects in the CT scan. Connected component labeling (CCL) is widely used for segmentation in CT baggage scans due to its efficiency. CCL first sets the foreground in the image and then uniquely labels subsets of connected components. Even though CCL can find foregrounds in the CT baggage scan, it is not enough to separate objects with similar intensities and deal with CT artifacts as it does not take account into intensity information. To prevent merging in CCL, we can apply morphological opening operation to CCL results. However, the morphological opening operation may remove thin sheet structures. Therefore, we find bulk objects in CCL result based on shape and apply morphological opening only for bulk objects.

To differentiate bulk objects from sheet objects, we extract shape features such as volume and surface area and feed them into supervised classifiers. In addition, we find the minimum volume enclosing ellipsoid [9] to extract shape features. After finding the minimum volume enclosing ellipsoid, we describe the shape as following:

- Ellipsoid Axes
- Axis ratio: minimum axis length / maximum axis length
- Volume ratio: object volume / ellipsoid volume

Given shape features, we find the classifier using support vector machine (SVM) [10]. SVM is a popular high-dimensional classifier in computer vision. SVM finds a hyper-plane that separates two groups by training the algorithm on a pre-classified set. From this pre-classified set, SVM selects a number of samples that are close to the opposite group. These samples are called support vectors and define the hyper-plane by maximizing margins between them. While the SVM classifier very effective in finding the hyper-plane, selecting the flexibility of hyper-plane is very important. If the hyper-plane is too stiff, then the classifier may not perfectly separate the two groups. On the other hand, if the hyper-plane is too flexible, then the classifier will over-train the data. We determine the type of hyper-plane by the non-linear Gaussian kernel that maps the data to a space where linear separation is possible. We automatically estimate the model parameter by 5-fold cross-validation on pre-classified set.

(a) (b) (c)

Figure 4. Effect of Selective Morphological Opening in ATR: (a) Input image, (b) CCL result followed by typical morphological opening, (c) CCL result followed by selective morphological opening on bulk objects. Notice that thin rubber sheet is preserved after morphological opening.

Figure 4 presents the benefit of our selective morphological opening in ATR. In the input image 4(a), there exist thin rubber sheet that is target to detect. Typical morphological opening erases the thin rubber sheet, as described in 4(b). Our selective morphological opening on bulk objects helps preserving thin structures while splitting merged bulk objects.

Metal Artifact Reduction in Image Domain

Streaking artifacts make CCL difficult to set the correct segmentation mask as streaking artifacts change the intensity. Therefore, we apply metal artifact reduction (MAR) on the image to correct intensity change due to streaking artifacts. MAR methods typically correct the metal trace in the raw CT data by using either sinogram inpainting [11] or the projection of a prior image that is derived from the original CT image [12]. The final image is then obtained by reconstructing the corrected sinogram. However, these sinogram-based MAR requires the raw CT data. Moreover, reconstructed image with the corrected sinogram may introduce secondary artifacts outside streaks. Therefore, we identify the metal artifact regions in the image domain and apply image-inpainting methods to generate the corrected image. In this way, MAR can be done in the image domain even though the original CT data is not available.

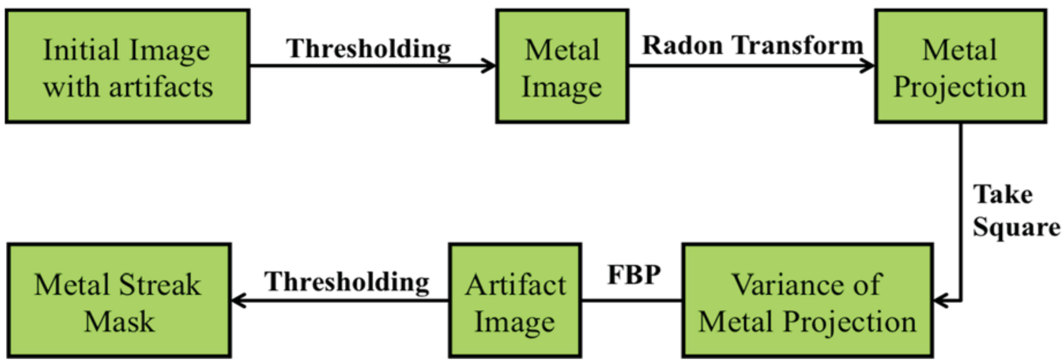
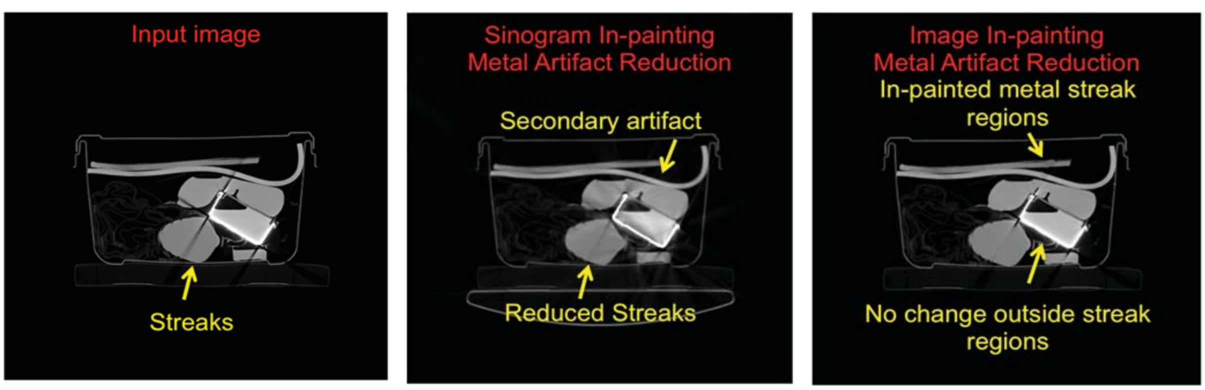


Figure 5. Flowchart to estimate variance of metal projection and detect streak mask in the image

Figure 5 illustrates the procedure of our metal artifact region detection. Given input image, we identify metal region through thresholding. Next, we forward-project the metal image via Radon transform and model non-linear variance of metal projection with square function. Finally, we filtered back-project the variance and threshold to generate metal streak mask. Here, we assume that X-ray through the metal introduces non-linear fluctuations to the projection measurements and the variance of projection measurements causes the streaking artifacts. This assumption is also made on Karimi's work at Task Order 3 [13, 14].



(a) (b) (c)

Figure 6. Comparison of MAR results: (a) Input image, (b) Sinogram Inpainting MAR, (c) Image Inpainting MAR. Notice that our image inpainting MAR does not introduce the secondary artifact outside the streak.

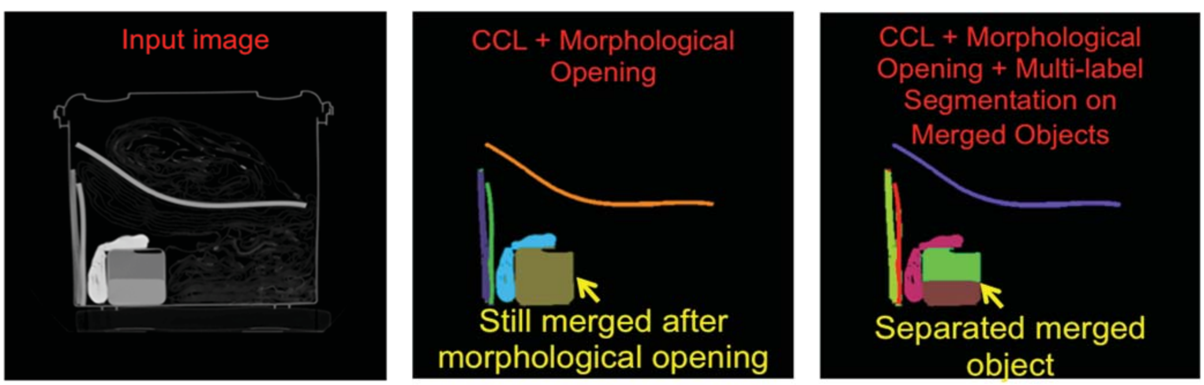
Figure 6 presents a sample of our image-inpainting MAR. In the input image 6(a), there exist metal streak. Sinogram inpainting MAR [15] reduces the streak artifact but introduces secondary artifact, as described in 6(b). Our image inpainting MAR only changes the intensity in the streak region preserving the resolution outside streak region.

Multi-label Segmentation on Merged Objects

Even though morphological opening can help split merged objects in CCL, it is not sufficient to separate highly cluttered objects due to tight packing because it does not take account into intensity information. For example, morphological opening cannot separate big merged objects with two different materials if morphological operator is small. Therefore, we further segment CCL results followed by morphological opening with the multi-label segmentation that utilizes intensity information.

Multi-label segmentation is well-studied computer vision techniques that partition the spatially continuous image domain into multiple regions with minimal total perimeter [16,17,18]. Mutli-label segmentation partitions the image domain into multiple disjoint sub-regions. Then, segmentation problem is modeled in the maximum a posteriori (MAP) framework that optimizes the combined cost function of the forward model and the prior model. The forward model represents the data fidelity such as intensity difference between pixel intensity and pre-determined label intensity. The prior model such as Markov Random Field (MRF) model regularizes the total perimeters of segmentation labels by enforcing the label smoothness.

In this multi-label segmentation, there are two important parameters. One is the number of materials and the other is mean intensity for materials. These parameters can be pre-determined. For example, we assume there are 3 target materials in the image and set the mean intensity empirically (i.e. Saline: 1050MHU, Rubber: 1200MHU, and Clay: 1500MHU). However, this pre-determined parameter would not be robust across images due to different level of CT noise and saline concentration, causing over-segmentation that splits the object incorrectly. So, we instead find the parameters adaptively based on peak analysis on histogram. For example, in this merged object, we find the 3 peaks in the histogram and the corresponding peak intensity. By using these adaptive parameters, we can achieve more accurate segmentation compared with using predetermined parameters.



(a) (b) (c)

Figure 7. Benefit of Multi-label Segmentation on Merged Objects: (a) Input image, (b) CCL result followed by our morphological opening, (c) After our muti-label segmentation on merged objects. Notice that our multi-label segmentation split the merged object based on intensity while preserving other objects without over-segmentation.

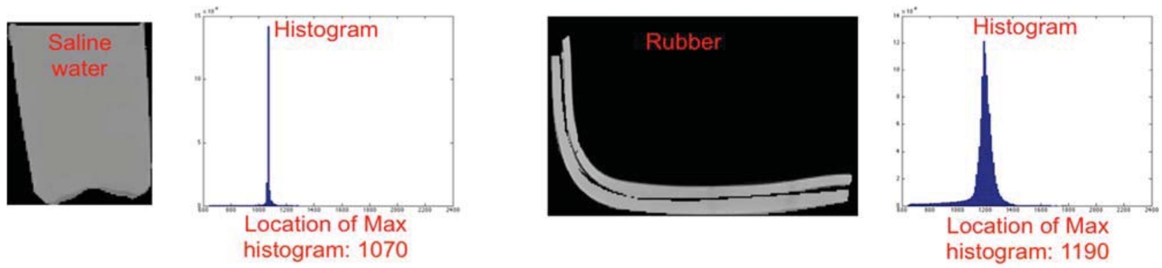
Figure 7 presents an example of benefit of our multi-label segmentation on merged objects in ATR. The CCL results after morphological opening in Figure 7(b) still include the merged object. As illustrated in the input image in Figure 7(a), this merged object contains two materials with significantly different intensities. Our multi-label segmentation on merged objects successfully split this merged object based on intensity while preserving other correctly segmented objects.

2. Feature Extraction

For each segmented object, we need to determine whether it is target or non-target. For this task, we first need to transform the image data into the set of features that describe the properties of the segmented object. Baseline ATR provided by ALERT uses the mass of each segmented object as feature. However, mass itself is not sufficient to describe the complex properties of targets. Therefore, we construct high-dimensional features, which are widely used in literatures:

- Intensity: Min, Max, Mean, Variance
- Physical: Mass
- Histogram: Location of max histogram, Normalized histogram

When performing analysis of complex data, one of the major problems originates from the number of features. Analysis of high-dimensional features generally requires a large amount of memory and computational time. Furthermore, parts of features can be redundant, which blurs the performance of analysis (curse of dimensionality). So, we apply feature-selection method called minimum-redundancy maximal-relevance (mRMR) [19] on high-dimensional features to find important features.



(a) (b)

Figure 8. Most important feature selected by mRMR: (a) saline water, (b) rubber; mRMR selects the location of max histogram as the most important feature. This reflects that material-specific CT number (attenuation coefficient) is the key of representing targets.

Figure 8 shows the most important feature selected by mRMR method among high-dimensional features. mRMR selects the location of max histogram as the most important feature reflecting that material-specific CT attenuation coefficient is the signature for target recognition. It is worth noting mRMR also selects the part of normalized histogram that enables us to capture subtle intensity difference between target and non-target.

3. Target Classification

Given extracted and selected features, we feed them into the classifier that determines whether the segmented object is target or not. Sample ATR provided by ALERT uses the binary threshold classifier for one-dimensional mass feature. Since our new features are multi-dimensional (after feature selection), we need the advanced classifier that can deal with high-dimensional features.

For target classification, we again use SVM classifier similar to sheet/bulk identifier but based on histogram features. Instead of finding one target classifier for all training objects, we first cluster the training set via K-mean algorithm [20]. K-mean algorithm finds K clusters in the observed data in a way that the sum of variances in the cluster is minimized. To define the variance in the cluster, we again use the shape features used for sheet/bulk identifier. By doing so, we can organize the dataset with similar shapes. We empirically set the number of cluster as 15 (i.e. K = 15) by balancing the number of targets and non-targets in each cluster to avoid over-training issues.

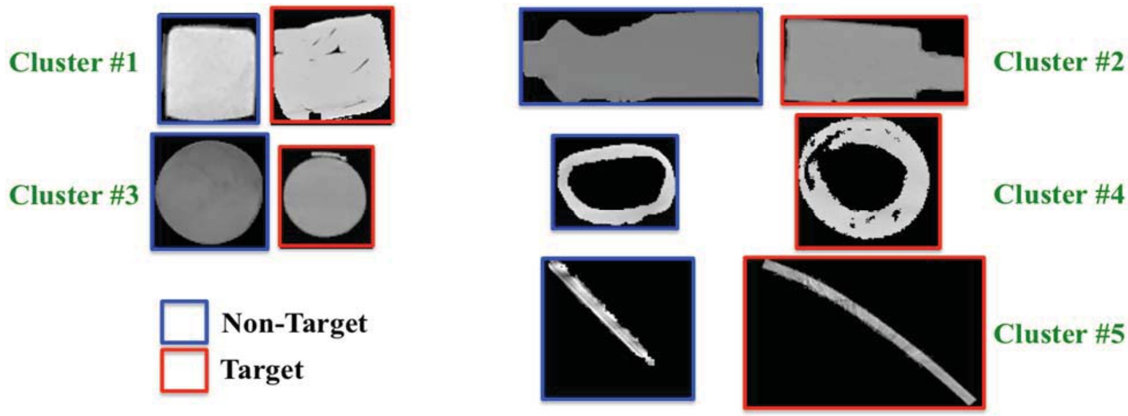
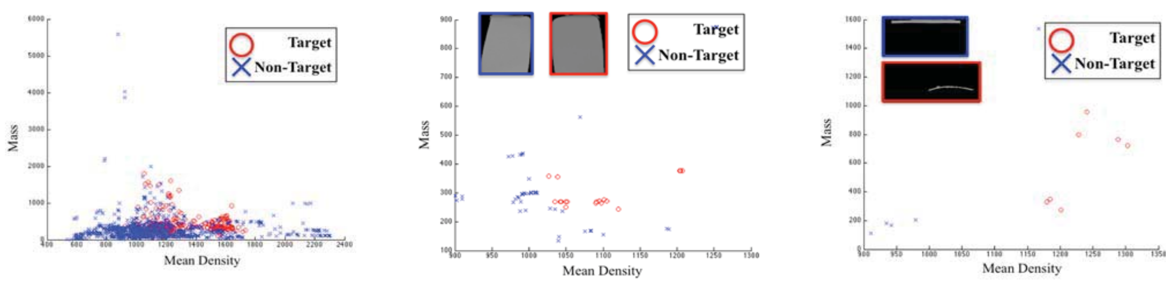


Figure 9. Examples of clusters on training dataset. Each cluster has the similar shape. Note the slight intensity difference between targets and non-targets in each cluster.

Figure 9 shows the examples of shape clusters determined by K-means algorithm. Blue and red rectangles represent the non-target and target, respectively. Then, each cluster has similar shape (i.e. cluster 1: square, cluster 2: bottle, cluster 3: circle, cluster 4: band, and cluster 5: sheet). In addition, we can notice that the intensity difference between target and non-target is slight in each cluster. Therefore, we can capture the slight intensity difference between targets and non-targets after this shape clustering.



(a) (b) (c)

Figure 10. Features in training set of: (a) all objects, (b) bottle shape cluster, (c) thin sheet shape cluster. Horizontal and vertical axes represent mean density and mass of each object, respectively. Notice that there is large overlap between target (red circle) and non-target (blue cross) for all training objects in (a). On the contrary, in each shape clusters in (b) and (c), features are more separable and thus it is easier to train the classifier.

Figure 10 illustrates how clustered training set can improve the classification performance. Figure 10 (a) shows the mean density and mass features of all training objects. It is worth noting that there is large overlap between target and non-target objects. The overlap in feature space makes it difficult to train the accurate classifier. Our shape clustering can help reducing the overlap in the feature space as described in Figure 10 (b) and (c). After shape clustering, features in each shape cluster are more separable and therefore it is easier to train the accurate classifier.

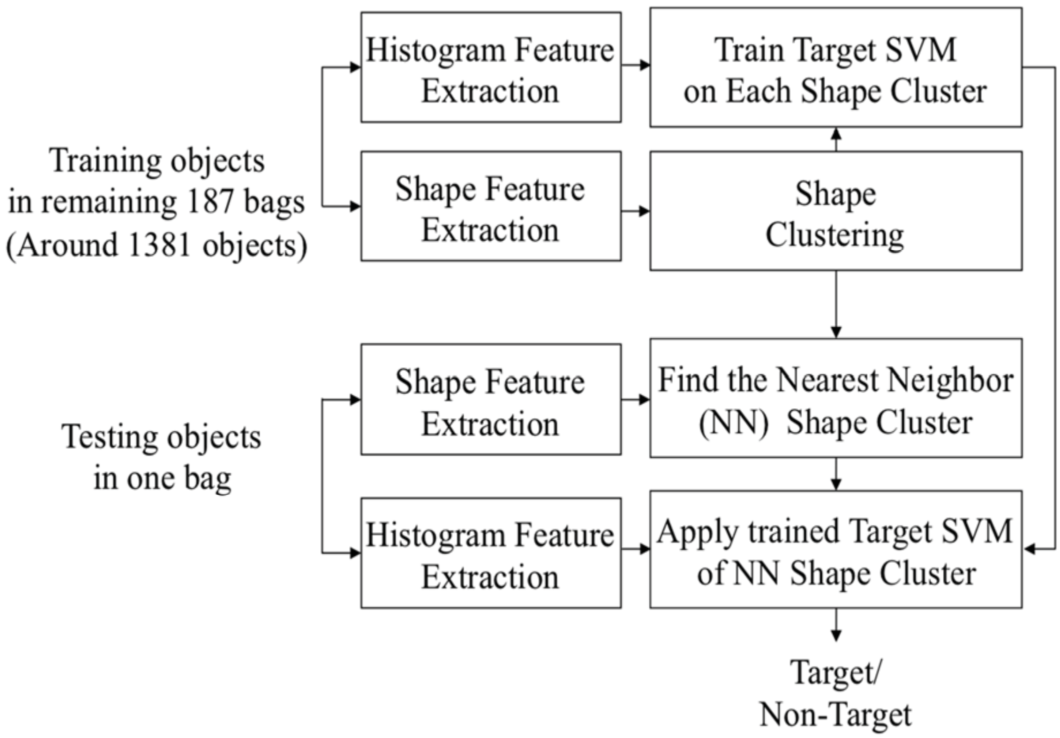


Figure 11. Flow-chart of target classification after shape clustering

After shape clustering, we perform the target classification. Figure 11 illustrates the flowchart of our target classification after shape clustering. For testing, we use leave-one-bag out cross validation scheme. We left out one bag for testing and train the target classifier with segmented objects in remaining 187 bags. Specifically, we extract shape features and perform clustering. For each shape cluster, we train the target SVM based on histogram features. For segmented objects in a testing bag, we extract shape features and find the nearest neighbor shape cluster in training set. We then feed histogram features to the SVM trained for nearest neighbor cluster to detect targets.

B. PD/PFA Results

Target Type	Target Subtype	Level of Difficulty	Num Targets	Num Detected	PD
Target	**All**	**All**	**407**	**387**	**0.95**
Target	Clay	All	111	106	0.96
Target	Rubber	All	158	151	0.96
Target	Saline	All	138	130	0.94
Target	Bulk	All	270	260	0.96
Target	Sheet	All	137	127	0.93
Target	All	Low	77	75	0.97
Target	Clay	Low	29	28	0.97
Target	Rubber	Low	22	21	0.96
Target	Saline	Low	26	26	1
Target	Bulk	Low	56	54	0.96
Target	Sheet	Low	21	21	1
Target	All	High	317	299	0.94
Target	Clay	High	82	78	0.95
Target	Rubber	High	125	119	0.95
Target	Saline	High	110	102	0.93
Target	Bulk	High	201	193	0.96
Target	Sheet	High	116	106	0.91
Pseudo-target	Sheet	High	10	9	0.9

Num Non-targets	Num FAs	PFA
1371	**110**	**0.08**

Table 1. Probability of Detection and Probability of False Alarms for all 188 scans

Table 1 shows Probability of Detection and Probability of False Alarms scores on all 188 scans with our ATR. We used nested 5-fold cross validation scheme to evaluate our ATR. In summary, we detected 387 targets among 407 total targets (0.95 Probability of Detection (PD)) and falsely detected 110 non-targets among 1371 non-targets (0.08 Probability of False Alarm (PFA)). As expected, we achieved better performance (PD: 0.97) for low difficulty cases than for high difficulty cases (PD: 0.94). Among high difficulty cases, PD for saline was lower (0.93) than that for clay and rubber (0.95). This is because it is challenging for our classifier to detect very low-density saline (<3.5% density). In terms of shape, PD for sheet was lower (0.91) than that for bulk (0.96). This is because our segmentation was not sufficient to prevent split in the thin structures by beam-hardening effects.

	CCL	CCL + Morphological Opening	CCL + Morphological Opening + Multi-label Segmentation	CCL + Morphological Opening + Metal Artifact Reduction + Multi-label Segmentation
PD	0.63	0.89	0.96	0.97

Table 2. Evaluation of Segmentation Methods

Table 2 shows PD after segmentation to indirectly measure the benefit of each component of our segmentation pipeline. Compared with baseline CCL provided by ALERT, our morphological opening significantly improves PD from 0.63 to 0.89. This indicates that our morphological opening better separates cluttered bulk objects while preserving thin objects. Multi-label segmentation further improves PD to 0.96. This reflects that our multi-label segmentation correctly splits the remaining merged objects based on intensity without over-segmentation issue. The gain from Metal Artifact Reduction is minor (from 0.96 to 0.97) in the segmentation. But, intensity correction from Metal Artifact Reduction will be critical to extract reliable features for target classifier.

	Our Segmentation Only	Our Segmentation + One Classifier for All Training	Our Segmentation + Clustered Target Classifiers
PD	0.97	0.89	0.95
PFA	0.73	0.16	0.08

Table 3. Evaluation of Classification Methods

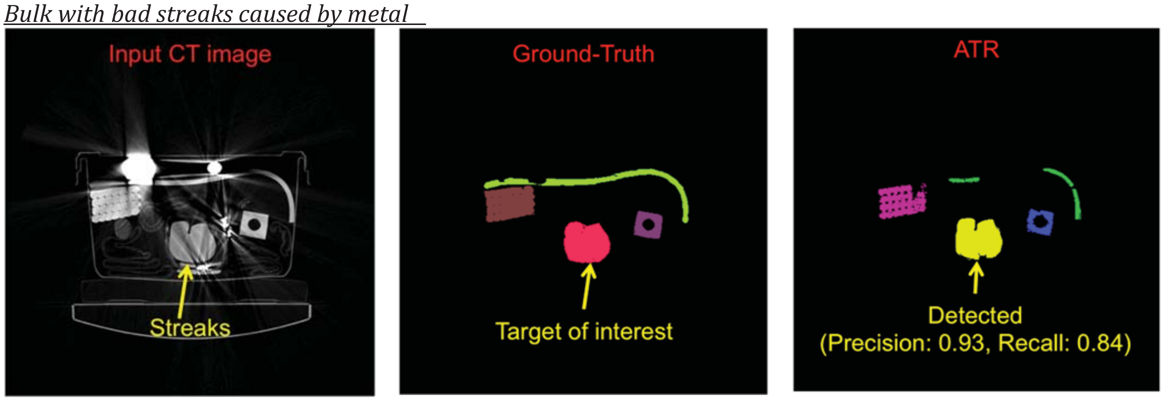
Table 3 shows PD/PFA to evaluate our classification pipeline. Without any classification, our segmentation achieves very high PD (0.97) but also very high PFA (0.73). This is natural, as we do not ignore non-targets from segmented objects. The classifier trained on all segmented objects decreases the PFA to 0.16 but significantly loses the target detection with 0.89 PD. This indicates that the classification performance is limited because the classifier is trained on highly overlapped features. On the contrary, our clustered target classifier significantly improves the classification performance with 0.95 PD and 0.08 PFA. This represents that clustered target classifier is critical to achieve higher PD and lower PFA in our ATR.

C. ATR Results

We will show our ATR results for 10 exemplar cases and 2 failure cases.

1. Example Cases

Bulk with bad streaks caused by metal

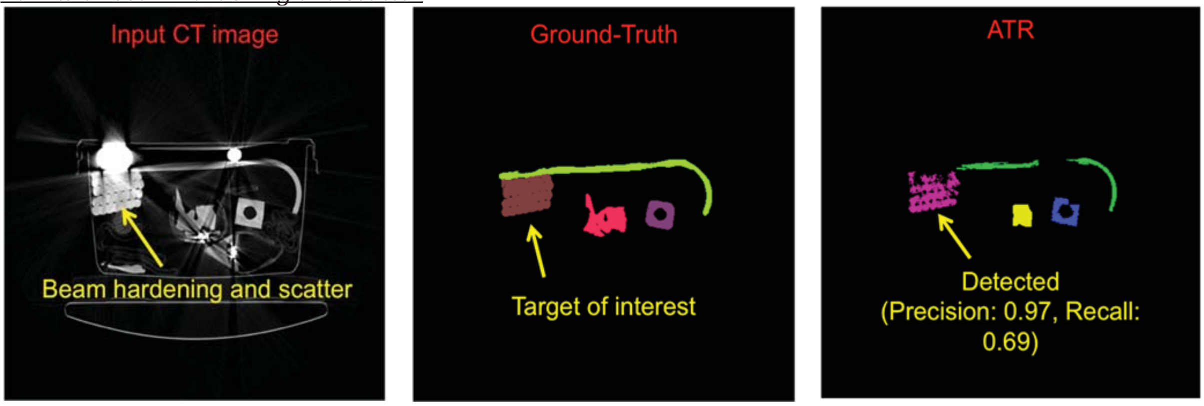


(a) (b) (c)

Figure 12. SSN 13 (Slice 105): (a) Input image, (b) Ground-Truth target, (c) Our ATR result. Notice that our ATR successfully detects the target with bad streak with a help of metal artifact reduction.

We show our ATR result with bulk object with bad streaks caused by metal in Figure 12. Thanks to our metal artifact reduction in image domain, the target is not split by bad streak and thus we can detect the target with high precision (0.93) and high recall (0.84).

Bulk with beam hardening and scatter

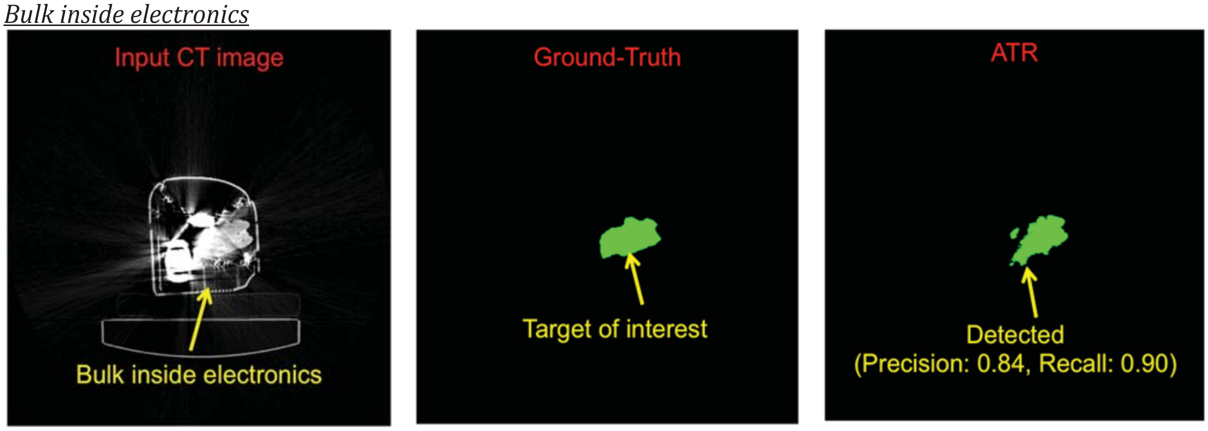


(a) (b) (c)

Figure 13. SSN 13 (Slice 128): (a) Input image, (b) Ground-Truth target, (c) Our ATR result. Notice that our ATR successfully detects the target with beam hardening and scatter.

We show our ATR result with bulk object with beam hardening and scatter in Figure 13. Since our ATR does not take account into beam hardening, the segmentation of target is partly affected, leading to relatively low recall rate (0.69). However, we use the important feature selected from the histogram and thus successfully detect the target.

Bulk inside electronics

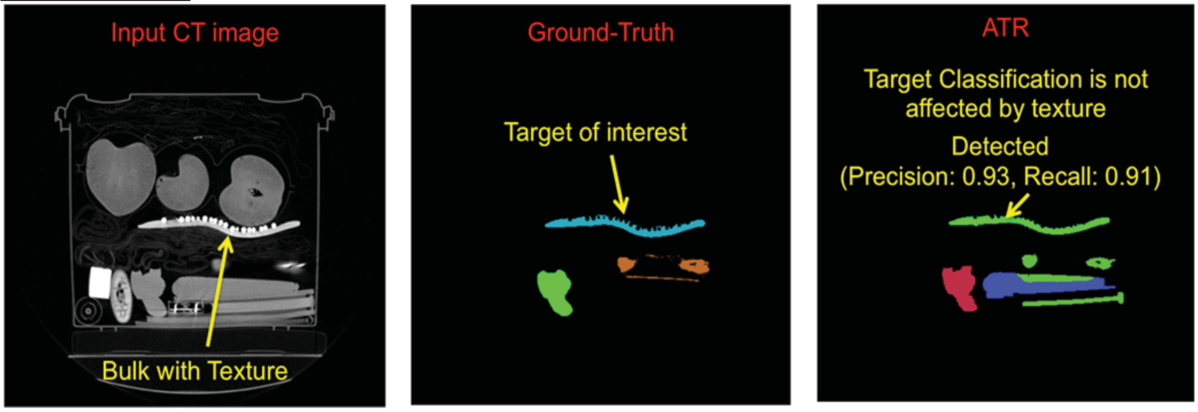


(a) (b) (c)

Figure 14. SSN 35 (Slice 49): (a) Input image, (b) Ground-Truth target, (c) Our ATR result. Notice that our ATR successfully detects the target with beam hardening and scatter.

We show our ATR result with bulk object inside electronics in Figure 14. Our merged object classifier prevents over-segmentation and preserves the correct mask from CCL. With a help of feature selection, our classifier correctly identifies target even with electronic noise.

Bulk with texture

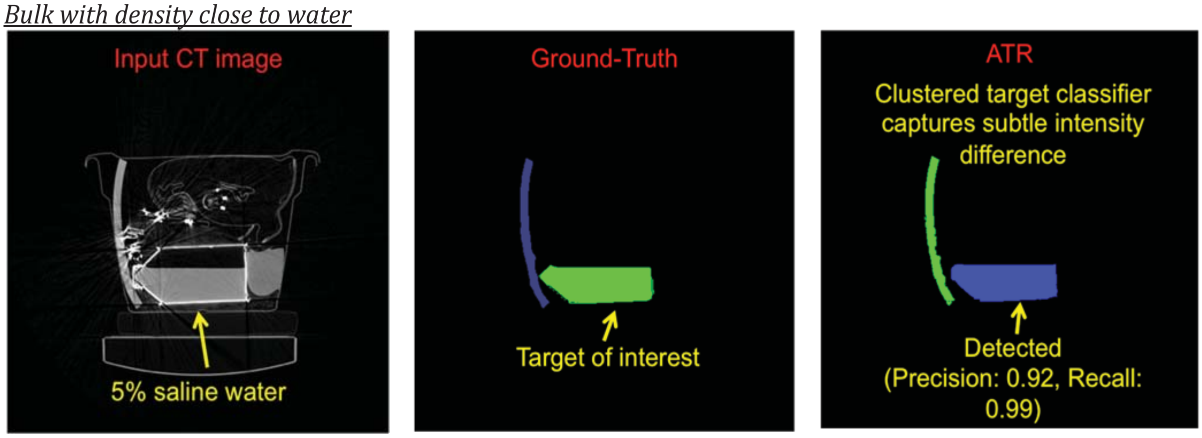


(a) (b) (c)

Figure 15. SSN 193 (Slice 198): (a) Input image, (b) Ground-Truth target, (c) Our ATR result. Notice that our ATR is not affected by texture.

We show our ATR result with bulk object with texture in Figure 15. Our segmentation uses the intensity information and thus preserves the texture of target of interest with high precision (0.93) and high recall (0.91). Then, our feature for target classifier is not based on texture but on histogram, our ATR did not ignore the target based on unique texture of this target.

Bulk with density close to water

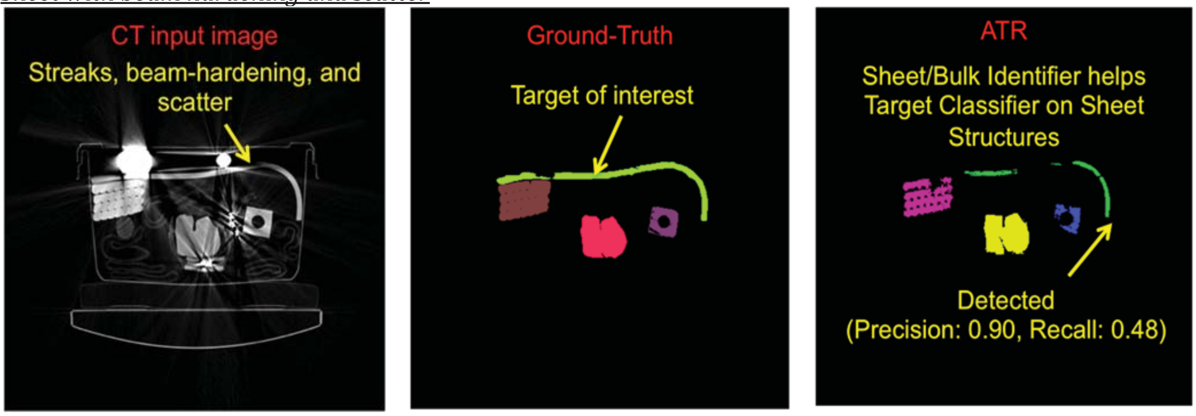


(a) (b) (c)

Figure 16. SSN 63 (Slice 45): (a) Input image, (b) Ground-Truth target, (c) Our ATR result. Notice that our ATR successfully detects the saline target whose density is close to water (5%).

We show our ATR result with bulk object with density close to water in Figure 16. It is worth noting that the intensity range of the target overlaps with that of water non-target because the density of saline is relatively low (5%). Our clustered target classifier captures those subtle intensity differences and successfully detects the target.

Sheet with beam hardening and scatter

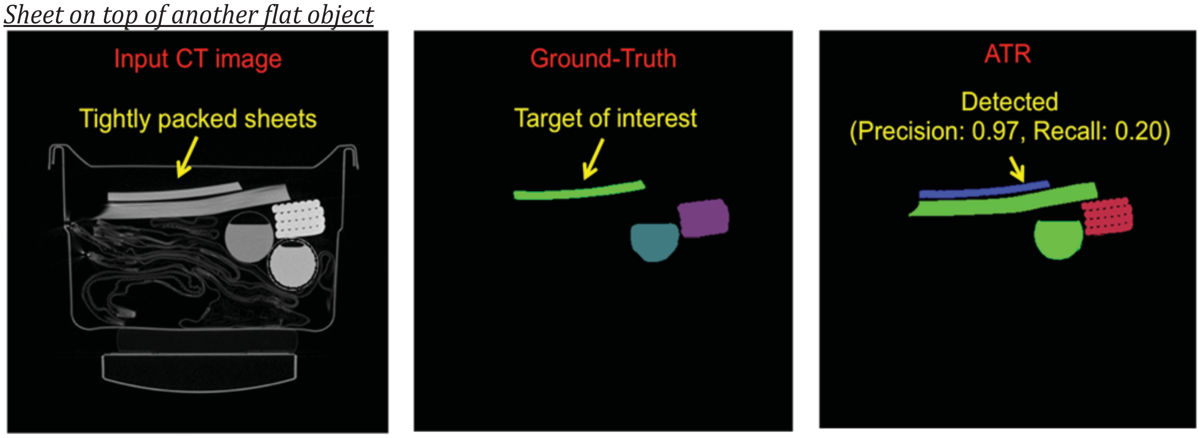


(a) (b) (c)

Figure 17. SSN 13 (Slice 111): (a) Input image, (b) Ground-Truth target, (c) Our ATR result. Notice that our ATR split the sheet by beam hardening and scatter but successfully detects the target.

We show our ATR result with sheet object with beam hardening and scatter in Figure 17. Since our ATR does not take account into beam hardening, the segmentation of target is partly affected, leading to relatively low recall rate (0.48). However, our sheet/bulk identifier prevents the sheet target from being erased by morphological opening and thus target classifier can detect the target

Sheet on top of another flat object



(a) (b) (c)

Figure 18. SSN 33 (Slice 46): (a) Input image, (b) Ground-Truth target, (c) Our ATR result. Notice that our ATR successfully splits tightly packed sheets.

We show our ATR result with sheet object on top of another flat object in Figure 18. Our multi-label segmentation on merged object can split tightly packed sheet object based on intensity information. Therefore, we can successfully detect the target.

Object with lots of photon starvation

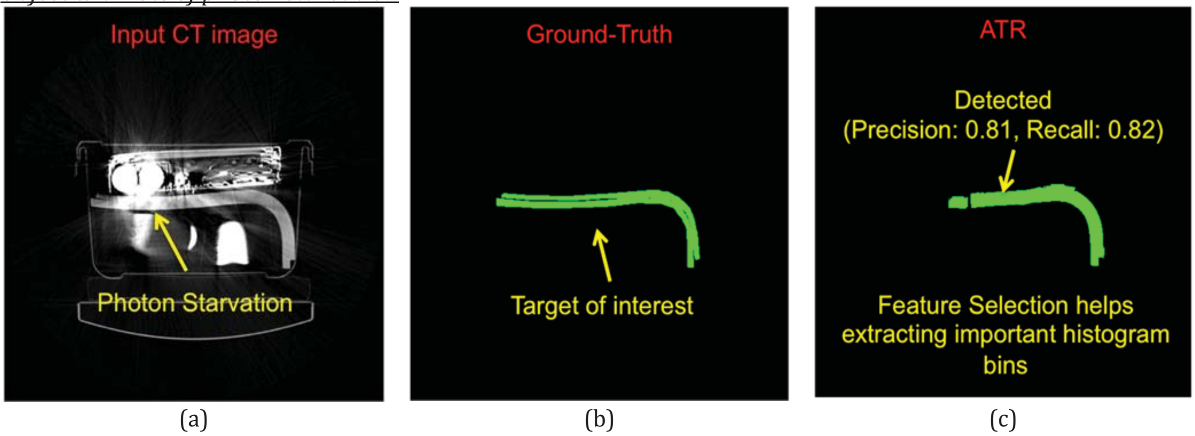


Figure 19. SSN 11 (Slice 94): (a) Input image, (b) Ground-Truth target, (c) Our ATR result. Notice that our ATR successfully detects target with lots of photon starvation thanks to feature selection.

We show our ATR result with the object with lots of photon starvation in Figure 19. Since our ATR does not take account into photon starvation, the segmentation is affected with loosing parts of the target. Then, feature selection helps extracting important histogram bins for target classifier to detect the target.

Pseudo-target sheet based on thickness

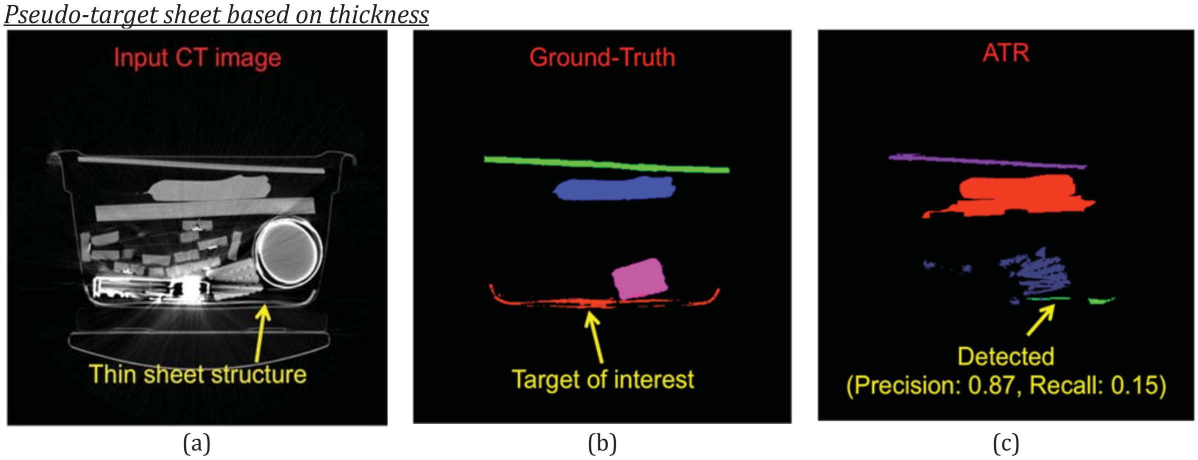


Figure 20. SSN 18 (Slice 125): (a) Input image, (b) Ground-Truth target, (c) Our ATR result. Notice that our ATR split the thin sheet by beam hardening but successfully detects the target.

We show our ATR result with the very thin pseudo-target in Figure 20. The segmentation is affected by beam hardening and we lose most of part of targets with low recall rate (0.15). But, the sheet/bulk identifier helps preserving this thin objects and classifier detects the target.

Pseudo-target powder based on density

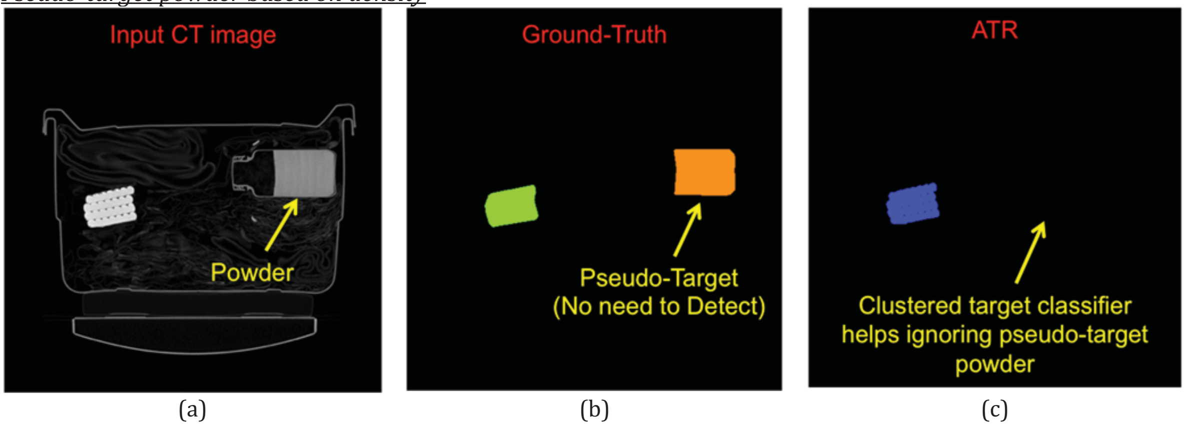


Figure 21. SSN 12 (Slice 105): (a) Input image, (b) Ground-Truth target, (c) Our ATR result. Notice that our ATR split the thin sheet by beam hardening but successfully detects the target.

We show our ATR result with the pseudo-target powder based on density in Figure 21. It is worth noting that we do not necessarily detect this pseudo-target powder. Even though our segmentation correctly identifies the pseudo-target powder, our clustered target classifier neglects this pseudo-target. This indicates that clustered target classifier is sensitive to subtle intensity difference in the similar shapes.

2. Failure Cases

Very low density saline

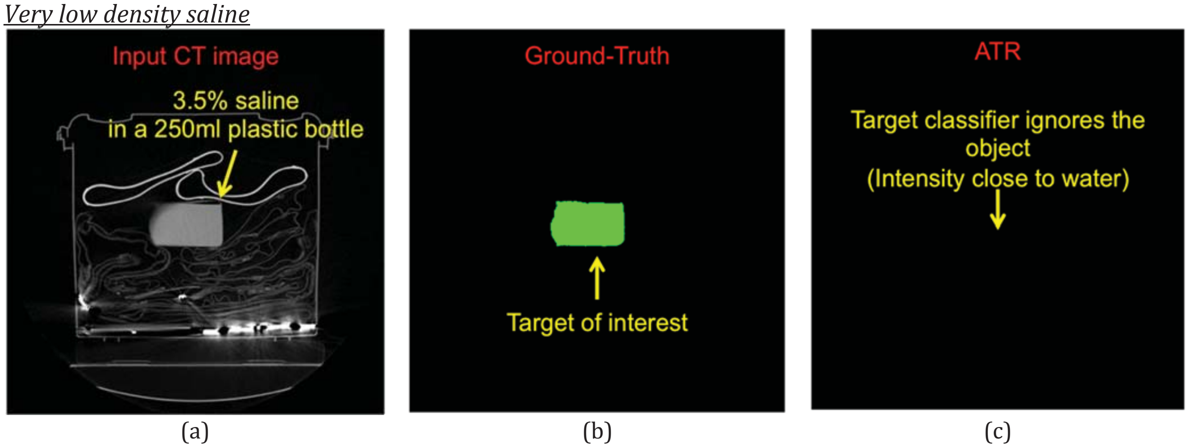
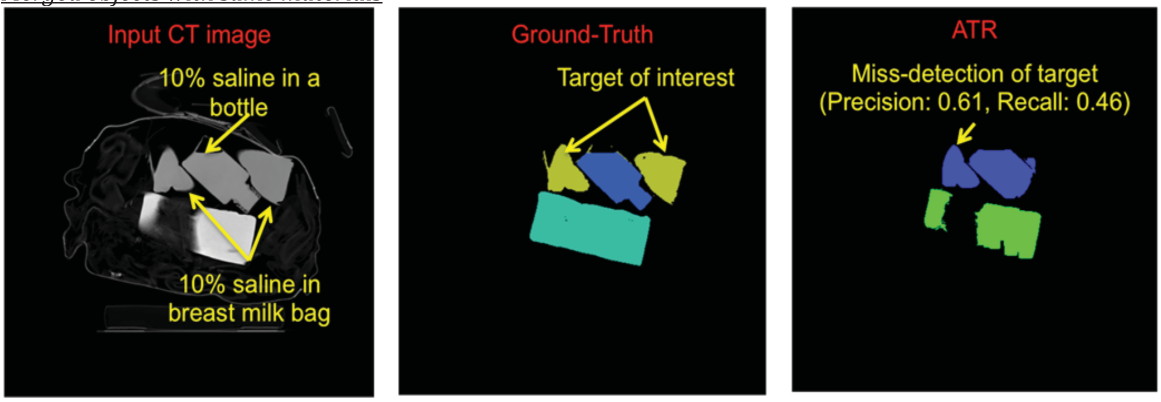


Figure 22. SSN 91 (Slice 125): (a) Input image, (b) Ground-Truth target, (c) Our ATR result. Notice that our ATR split the thin sheet by beam hardening but successfully detects the target.

We show our ATR result with very low-density saline target in Figure 22. In the input image, there is 3.5% saline target in a 250ml plastic bottle. Since the low-density saline is in the small plastic container, the intensity range of the target is very close to that of water. Therefore, our ATR ignores the target even though the segmentation identifies the object.

<u>*Merged objects with same materials*</u>



(a) (b) (c)

Figure 23. SSN 17 (Slice 50): (a) Input image, (b) Ground-Truth target, (c) Our ATR result. Notice that our ATR split the thin sheet by beam hardening but successfully detects the target.

We show our ATR result with merged objects with same materials in Figure 23. In the input CT image, we note that one 10% saline in a bottle and one 10% saline in breast milk bag are closely located. Since two objects have the same material (10% saline), intensity level is almost identical. Then, our multi-label segmentation on merged objects gives incorrect segmentation, as it cannot capture the intensity difference between two objects. This miss-segmentation leads to miss-detection of target with low recall rate (0.46).

D. *Major Contributions*

We have developed and implemented an ATR system for CT baggage scans, adapted the advanced image segmentation, feature extraction, and target classification in computer vision to a particular security screening CT systems. The major contributions are listed in the following:

- Image Segmentation
  - Multi-label Segmentation on Merged Objects
- Feature Extraction
  - Feature Selection From High-dimensional Histogram Bins
- Target Classification
  - Shape Clustered Classifier

We meet the goal of project in terms of PD (95%) and PFA (8%). The proposed segmentation algorithm splits the merged objects by tight packing and corrects the partial loss due to the metal artifacts. Our classification helps differentiating targets from non-targets decreasing false alarms. These are all factors that can lead to improved target detection in ATR systems.

E. *Future plans*

In the near term, we will continue to refine our ATR system to produce higher probability of detection and lower probability of false alarm. To do this, we will investigate advanced features and classifiers to decrease the probability of false alarm while not ignoring targets. For example, we will try Gabor features that model the behavior of cells in the visual cortex and Deep Learning classifier that models the neuron network in the brain. Once this task is complete, we will identify the most challenging cases (corner case) including severe beam hardening artifacts or systematic errors in CT number and formulate improved forward and prior models in our segmentation algorithm in order to enhance segmentation quality. We hope that this will ultimately improve the probability of detection while reducing the false alarm rates.

IV. RELEVANCE AND TRANSITION

Our transition task was, by its very design, highly relevant to the goals of DHS in involving people outside the vendors and improving detection of threats in checked baggage while minimizing the cost related secondary inspection. We have shown that our ATR provides the sort of quality improvement. The segmentation/ classification dealing with merged/divided objects by tight packing/metal artifacts shown above is a potential gain, and we hope that this will translate into automated detection performance.

V. PROJECT DOCUMENTATION AND DELIVERABLES

A. Peer Reviewed Conference Proceedings

1. P. Jin, D. Ye and C. Bouman, "Joint Metal Artifact Reduction and Segmentation of CT Images using Learning Dictionary-Based Image Prior and Continuous-relaxed Potts Model", to be submitted to *IEEE International Conference on Image Processing*, Quebec City, Canada, 2015.

B. Transferred Technology/Patents

N/A

C. Software Developed

Our principal focus has been algorithm development in implementation of ATR for the CT baggage scan screening. Transition into application is envisioned with subsequent work on computational efficiency.

VI. BIOGRAPHIES

Dong Hye Ye is a Postdoctoral Researcher in Electrical and Computer Engineering at the Purdue University. His research interests are in advancing machine learning for image processing. His publications have been awarded Best Paper at MICCAI-MedIA 2010 and Student Travel Grants at ISBI 2012, and PRNI 2012. During his PhD, Dong Hye conducted research at Section of Biomedical Image Analysis (SBIA) in Hospital of the University of Pennsylvania (HUP) and Microsoft Research Cambridge (MSRC). He received Bachelor's degree from Seoul National University in 2007 and Master's degree from Georgia Institute of Technology in 2008.

Pengchong Jin received his B.Eng. degree in Electronic and Computer Engineering with the first class honor from The Hong Kong University of Science and Technology in 2009, and is currently pursing the Ph.D. degree in School of Electrical and Computer Engineering, Purdue University. His research interests are in statistical signal and image processing, inverse problems, computer vision and machine learning.

Charles A. Bouman is the Showalter Professor of Electrical and Computer Engineering and Biomedical Engineering at Purdue University where he also serves has a co-director of Purdue's Magnetic Resonance Imaging Facility. He received his B.S.E.E. degree from the University of Pennsylvania, M.S. degree from the University of California at Berkeley, and Ph.D. from Princeton University in 1989. Professor Bouman's research focuses on inverse problems, stochastic modeling, and their application in a wide variety of imaging problems including tomographic reconstruction and image processing and rendering. Prof. Bouman is a Fellow of the IEEE, AIMBE, IS&T, and SPIE and is currently the IEEE Signal Processing Society's Vice President of Technical Directions. He has also served as the Editor-in-Chief of the IEEE Transactions on Image Processing and the Vice President of Publications for the IS&T Society.

VII. REFERENCES

[1] K. D. Sauer and C. A. Bouman, "A local update strategy for iterative reconstruction from projections," *IEEE Trans. on Signal Processing*, vol. 41, no 2, 1993
[2] C. A. Bouman and K. D. Sauer, "A Unified Approach to Statistical Tomography using Coordinate Descent Optimization," *IEEE Trans. on Image Processing*, vol. 5, no. 3, pp. 480-492, March 1996.
[3] C. Bouman and K. Sauer, "A Generalized Gaussian Image Model for Edge-Preserving MAP Estimation," *IEEE Trans. on Image Processing*, vol. 2, no. 3, pp. 296-310, July 1993.

[4] J. B. Thibault, K. D. Sauer, C. A. Bouman, and J. Hsieh, “A Three-Dimensional Statistical Approach to Improved Image Quality for Multi-Slice Helical CT,” *Med. Phys.,* vol. 34, no. 11, pp. 4526-4544, 2007.

[5] Z. Ying, R. Naiu and C. Crawford, “Dual Energy Computed Tomography for Explosive Detection,” Journal of X-ray Science and Technology, vol. 14, 2006, pp. 235-256

[6] L. Grady, V. Singh, and T. Kohlberger, “Automatic Segmentation of Unknown Objects, with Application to Baggage Security”, European Conference Computer Vision, pp. 430-444, 2012

[7] S. Singh and M. Singh, “Explosive detection systems (EDS) for aviation security”, Signal Processing 83 pp. 31-55, 2003

[8] C. Crawford, Algorithm Development for Security Applications Workshop 02, 2009

[9] P. Kumar and E. A. Yildirim, “Minimum-Volume Enclosing Ellipsoids and Core Sets”, Journal of Optimization Theory and Applications 126, pp.1-21, 2005

[10] C. Cortes and V. Vapnik, “Support-vector networks”, Machine Learning, vol. 20, pp. 273-297, 1995

[11] Y. Kim, S. Yoon, and J. Yi, “Effective sinogram-inpainting for metal artifact reduction in X-ray CT images”, IEEE International Conference Image Processing, pp.597-600, 2010

[12] E. Meyer, R. Raupach, M. Lell, B. Schmidt, and M. Kachelriess, “Normalized metal artifact reduction (NMAR) in computed tomography”, Medical Physics, vol. 37, no. 10, pp. 5482-5493, Oct. 2010

[13] S. Karimi, P. Cosman, and H. Martz, “Metal Artifact Reduction For CT-Based Luggage Screening”, IEEE International Conference on Acoustic, Speech and Signal Processing, pp. 1170-1174, 2014

[14] S. Karimi, Algorithm Development for Security Application Workshop 09, 2013

[15] H. K. Tuy, “A post-processing algorithm to reduce metallic clip artifacts in CT images”, European Radiology, vol. 3, no. 2, pp. 129-134, 1993

[16] J. Lellmann, F. Becker, and C. Schnorr, “Convex optimization for multi-class image labeling with a novel family of total variation based regularizers”, International Conference on Computer Vision, pp. 646-653, 2009

[17] Jing Yuan, Egil Bae, Xue-Cheng Tai, and Yuri Boykov, “A Fast Continuous Max-Flow Approach to Potts Model”, European Conference Computer Vision, 2012

[18] Nikos Komodakis and Georgios Tziritas, “Approximate labeling via graph-cuts based on linear programming”, IEEE Transactions on Pattern Analysis and Machine Intelligence, pp. 1436-1453, 2007

[19] H. Peng, F. Long, and C. Ding, “Feature selection based on mutual information: Criteria of max-dependency, max-relevance, and min-redundancy”, Pattern Analysis and Machine Intelligence, pp. 1226-1238, 2005

[20] E. Forgy, “Cluster analysis of multivariate data: efficiency versus interpretability of classifications”, Biometrics 21, pp. 768-769, 1965

11.6.2 ATR Development – Gregor

"ALERT Task Order 4: Final Report ATR Development Using Graph Algorithms"

ALERT Task Order 4: Final Report ATR Development Using Graph Algorithms

Dr. Jens Gregor
Dept. Electrical Engineering & Computer Science
University of Tennessee, Knoxville, TN 37996-2250

jgregor@eecs.utk.edu

January 21, 2015

Executive Summary

This report describes the development of an ATR (Automated Target Recognition) system for the separation of objects found in luggage into targets and non-targets. A volumetric X-ray CT image is first segmented into objects and background. Numerical features are then extracted from the objects and processed by a classifier, which has been trained to distinguish targets from non-targets.

The work presented is centered around segmentation based on graph algorithms. The data flow is similar to prior art in that segmentation produces objects that are classified into targets and non-targets using numerical features extracted from the objects, and that the segmentation consists of merging and splitting of objects produced by an initial labeling step. The segmentation part of the ATR differs from the prior art known to the present author by being based on graph models of object shape combined with Gaussian mixture models of object density distributions. Classification uses a support vector machine (SVM) framework applied to features representing object size and density statistics. The main contribution for the classifier part of the ATR consists of a k-nearest neighbor (kNN) based editing scheme applied during the parameter estimation phase.

Performance wise, the following PD rates were obtained: 85% for bulk targets, 95% for sheet targets, and 70% for sheet pseudo-targets. PFA came in at 20%. The conjecture is that a reduction of the PFA requires use of more features, which in turns calls for an improvement in the quality and consistency of the density distributions for the target objects. One possible course of action might be to have the graph algorithms that split merged objects based on their shape take density information into consideration and likewise have the Gaussian mixture model based density splitting algorithm account for object shape. Currently, the two operate without knowledge of the other. The goal would be to improve the segmentation accuracy beyond what is needed to meet the given recall and precision requirements. Another possible course of action would be to improve the overall image quality, e.g. by means of metal artifact reduction which was not applied here. This would have a direct impact on the shape and the extent of many of the object density distributions.

1. Introduction

1.1 Problem Specification and Main Results

This report describes the development of an ATR (Automated Target Recognition) system and its application to approximately 200 luggage scans obtained using an Imatron C300 X-ray medical CT scanner. The targets consisted of 407 saline, rubber and clay bulk and sheet objects as well as 10 sheet pseudo targets which were thinner and/or had lower mass than the ordinary targets.

The TO4 management team specified the performance goal to be better than 90% probability of detection (PD) for all bulk and sheet targets as well as sheet pseudo-targets, along with less than 10% probability of false alarm (PFA). Tables 1 and 2 summarize the results achieved. The developed ATR came close to performing as desired with a PD of 88.2% for targets overall, 84.8% for bulk targets, 94.9% for sheet targets, and 70% for sheet pseudo-targets. PFA came in at 20%. We address these results in greater detail below.

The data flow of the ATR is similar to prior art, e.g. (Kwon et al, 2013), in that segmentation produces objects that are classified into targets and non-targets and that the segmentation consists of merging and splitting of objects produced by an initial labeling step. The segmentation part of the ATR differs from the prior art known by the present author by being based on graph models of object shape combined with Gaussian mixture models of object density distributions. Using features that represent object size and density statistics, classification is based on a support vector machine framework (Cortes and Vapnik, 1995). The main contribution for the classifier part of the ATR consists of a k-nearest neighbor (kNN) based editing scheme applied during the parameter estimation phase.

1.2 ATR Design Philosophy and Overview

The ATR was designed with the following philosophy in mind. High PD requires high recall and precision, which translates into a need for accurate segmentation of the objects. Low PFA can be achieved by subsequent classification of objects as targets and non-targets using numerical features that describe object size and density characteristics. The Principle of Occam's Razor, which states that the hypothesis with the fewest assumptions should be selected, was applied in the sense that simplicity was used as a governing design principle through out.

The flow-chart shown in Figure 1 outlines the four main steps of the ATR, namely, connected component labeling based extraction of objects from the volumetric X-ray CT images, graph based splitting of merged objects with respect to their shape, Gaussian mixture model based splitting of merged objects with respect to their density distributions, and numerical feature extraction, support vector machine training and object classification. Each of these steps is described in greater detail below.

Table 1. Probability of Detection (PD) Results.

Target Type	Target Subtype	Level of Difficulty	Num Targets	Num Detected	PD [%]
Target	**All**	**All**	**407**	**359**	**88.2**
Target	Clay	All	111	95	85.6
Target	Rubber	All	158	148	93.7
Target	Saline	All	138	116	84.1
Target	**Bulk**	**All**	**270**	**229**	**84.8**
Target	**Sheet**	**All**	**137**	**130**	**94.9**
Target	All	Low	77	68	88.3
Target	Clay	Low	29	26	89.7
Target	Rubber	Low	22	21	95.5
Target	Saline	Low	26	21	80.8
Target	Bulk	Low	56	47	83.9
Target	Sheet	Low	21	21	100
Target	All	High	317	279	88
Target	Clay	High	82	69	84.1
Target	Rubber	High	125	117	93.6
Target	Saline	High	110	93	84.5
Target	Bulk	High	201	170	84.6
Target	Sheet	High	116	109	94
Pseudo Target	**Sheet**	**High**	**10**	**7**	**70**

Table 2. Probability of False Alarm (PFA) Results.

Num Non-targets	Num FAs	PFA [%]
1371	274	20

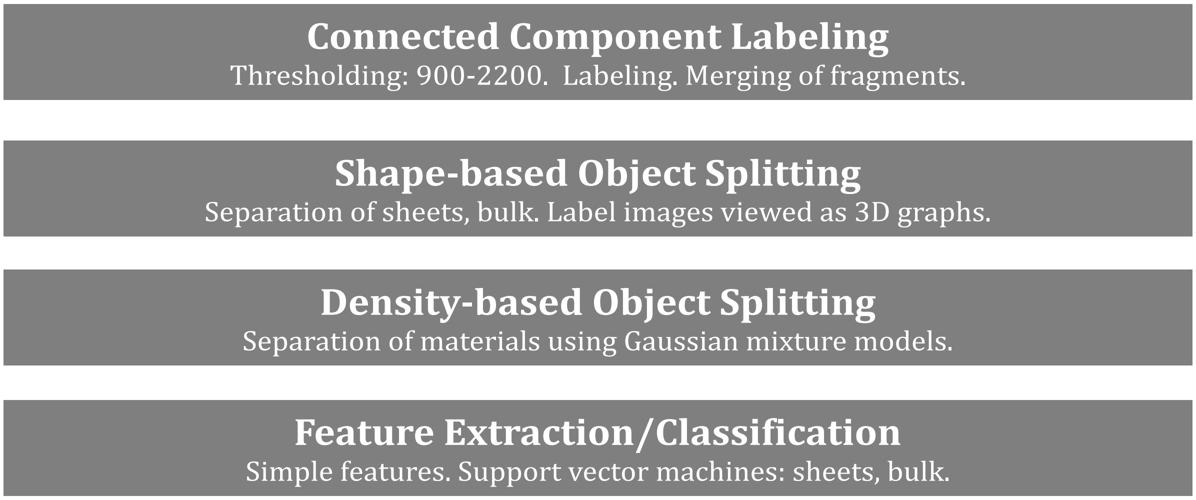


Figure 1. ATR flow-chart. The four main steps consist of i) connected component labeling based object extraction, ii) graph based object splitting with respect to shape, iii) Gaussian mixture model based splitting with respect to density, and iv) feature extraction and support vector machine training followed by classification of objects into targets and non-targets.

Connected component labeling (Gonzalez and Woods, 2008) was used to assign initial object labels to voxels in the density range 900-2200 which covers the range associated with the target materials under normal imaging conditions (that is, when not distorted by metal streak and shading artifacts). Connectivity was defined as a density difference of 75 or less between first-order neighbors. For voxel (i,j,k) the latter translates into the voxels given by (i,j,k)±(1,1,1). Morphological opening and closing was applied prior to object extraction in order to eliminate small structures and fill in small gaps. Post extraction, all objects were temporarily subjected to morphological dilation. Overlap regions resulting from this operation were used to determine pairs of objects that were at most one voxel apart. Any such adjacent objects were merged subject to the following simple rule regarding density mean and standard deviation statistics:

$$|\mu_a - \mu_b| < 2(\sigma_a + \sigma_b)$$

The goal of the object merger was to limit fragmentation due to streak artifacts and other density discontinuities.

Figure 2. The connected component based segmentation and the subsequent merging of object fragments may result in multiple objects being assigned the same label. Such merged objects are subjected to shape and density based splitting.

The connected component labeling and the subsequent merging of object fragments both operate without knowledge of object shape and density characteristics. As a result, objects may become fused in the process. As illustrated by Figure 2, the required subsequent splitting was carried out first with respect to shape and then density.

The shape based splitting was centered around graph models of the object label images. Graphs are abstract data structures, which can be used to model spatial relations (Cormen et al, 1986). Breadth-first and depth-first search algorithms were developed by the author to support separation of sheets from bulk objects followed by separation of two or more bulk objects; separation of sheets that lay on top of one another was not considered. The density based splitting was based on creating Gaussian mixture models (Press et al, 2007) of histograms representing the object density distributions and applying a Bayesian error minimization scheme to determine a possible separation threshold (Duda et al, 2001). We provide more detailed descriptions of these two approaches in Sections 2 and 3.

Following extraction of numerical features that describe object size and density statistics, two support vector machines (Cortes and Vapnik, 1995) were trained to classify sheet and bulk objects, respectively. The distinction between the two types of objects was done automatically using information extracted from the graph models of the object label images. The distinction between target and non-target training samples was also done automatically. The geometric mean of the recall and precision was computed for each object by comparing it against the set of known targets. If the largest such value exceeded a given threshold for the object type, namely, 25% for bulk targets, 10% for sheet targets, and 5% for sheet pseudo-targets corresponding to the given recall and precision thresholds x 0.50, the object was labeled as a target and used as such during training of the classifier. The idea behind this approach was to handle cases of imperfect segmentation for which it was impossible to uniquely label an object as target or non-target. The alternative of using the ground truth masks to extract the target features directly was not considered as it would have implied by-passing of the segmentation part of the ATR during classifier training. We elaborate on classification in Section 4.

2 Graph Based Object Splitting

Each object label image can be viewed as a graph with the voxels representing the vertices and the first-order neighborhood used for the connected component labeling defining the edges. Note that the graph model is an abstraction in that a graph is never actually formed.

Graph based splitting of merged objects is based on computing a distance map that describes the distance from the surface to the core (innermost voxels) of an object. Breadth-first search is a traversal algorithm for systematically visiting all vertices from a given start vertex using a queue (Cormen et al, 1986). We have developed a modified breadth-first search algorithm that computes the aforementioned distance map as follows. The set of surface voxels (those that have one or more non-object neighbors) are initialized to have a distance of 1 with all other voxels set to have distance of infinity. A breadth-first search is then carried out to propagate distances from the former to the latter, one layer of voxels at a time. Figure 3a illustrates the process and the end result. Bulk objects are separated from one another by determining the number and location of distinct core regions (sets of near-core voxels that form connected components) and assigning them new, unique object labels. By backtracking the distance map computation, these labels are propagated from the core regions back out to the surface, again one layer of voxels at a time. The backpropagation is halted either when a voxel would be assigned multiple labels or when the surface is reached. Figure 3b illustrates this process and the end result. Prior to separation of fused bulk objects, sheets are detected and split off as described next. Sheets differ from bulk objects by containing low-distance local maxima core voxels from which the surface can be reached in two opposing directions. These are detected using depth-first search, which is a traversal algorithm for finding a path from a source to a sink vertex using a stack (Cormen et al, 1986). See Figure 3c for an illustration of this process and the end result.

Figure 4 illustrates the graph based splitting of sheets from bulk objects. Surface voxels are colored a bluish purple in the distance maps while core voxels are colored yellow or white depending on their depth location. The resulting label images show that separation was achieved. However, we note that the bulk objects all bleed into the sheets. We speculate that use of density statistics may help reduce or even eliminate this tendency.

Figure 5 illustrates the graph based splitting of multiple bulk objects. Seven core regions are detected for the case in the upper panel. Two core regions are detected for the case in the lower panel. The resulting label images show that separation was achieved. Some degree of bleeding of one object into another is evident here as well. We again speculate that use of density statistics may help reduce or even eliminate this tendency. We note that separation was made possible by the concavities seen between the objects since without these, multiple core regions would not be detected during the initial phase of the computation. Flat objects stacked on top of each other are thus not separable by this method, be that sheets or bulk objects.

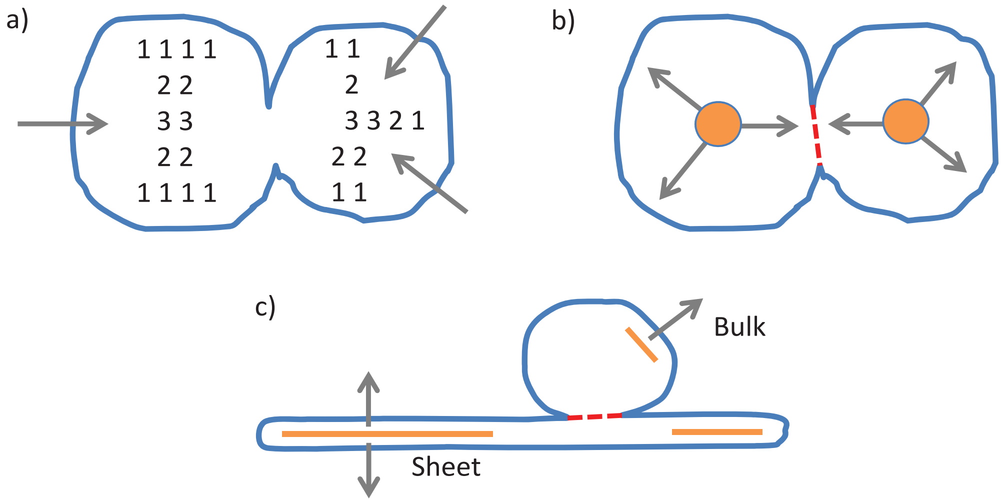


Figure 3. (a) Breadth-first search is used to compute a distance map that describes the distance from an object's surface to its core. (b) Retracing from selected core regions allows for splitting into multiple objects. (c) Depth-first search facilitates detection and separation of sheets. Sheets differ from bulk objects by having the surface be reachable in two opposing directions compared with the one direction in which a bulk surface is found.

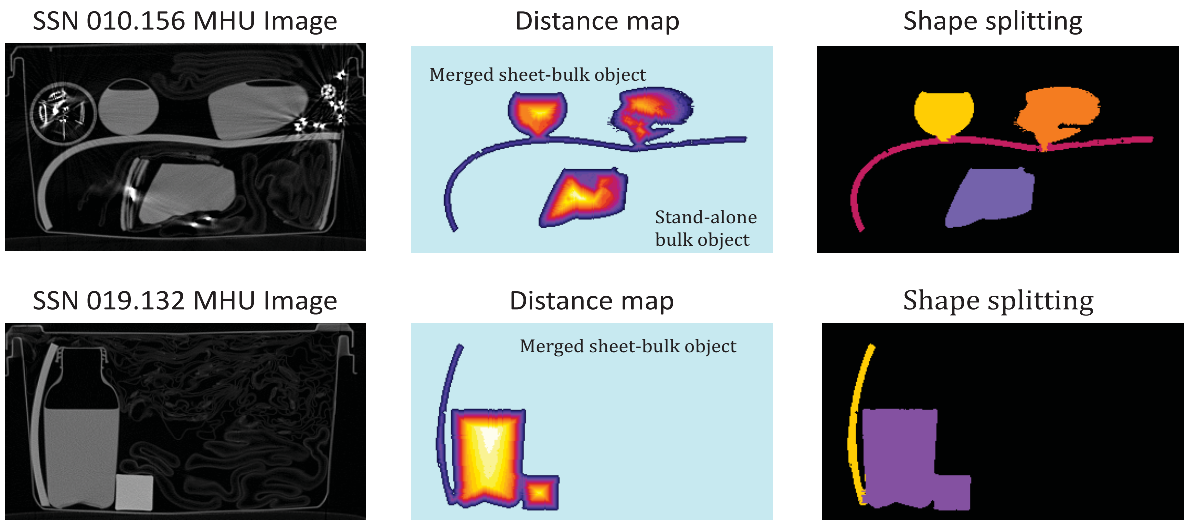


Figure 4. Graph based splitting of merged sheet and bulk objects.

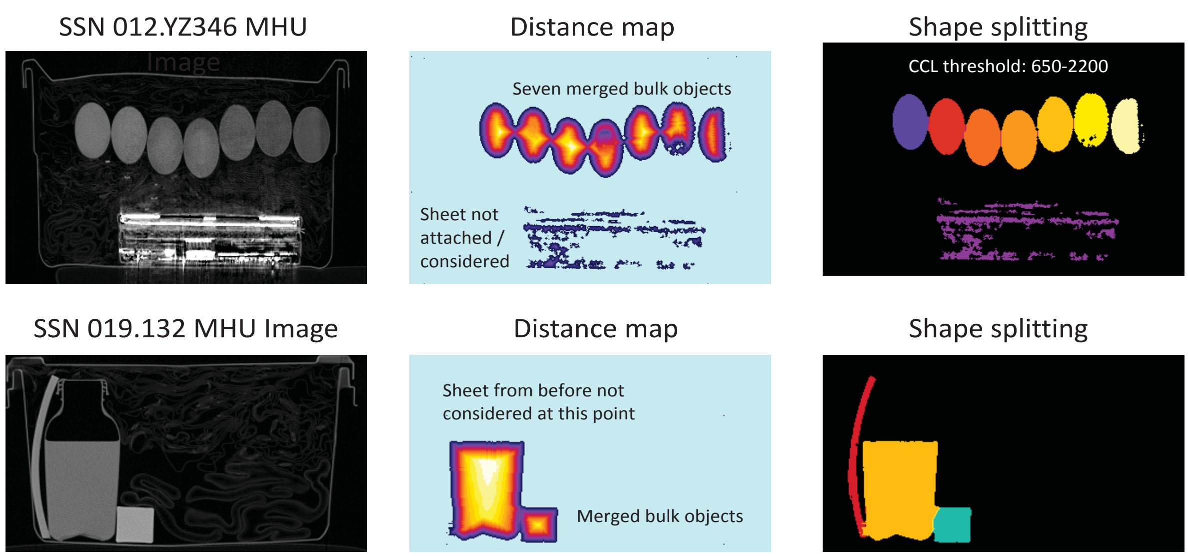


Figure 5. Graph based splitting of two or more merged bulk objects.

3 Gaussian Mixture Model Based Splitting

The density histogram for an object is modeled as a mixture (sum) of a fixed number (K) of Gaussian distributions. Expectation-maximization is used to estimate the underlying distribution parameters (Press et al, 2007). We used models having K=1, 2, and 3 mixture components. Figure 6 gives an example of K=2 versus K=3. A Chi-squared like test is used to determine the best fit. This fit is computed for all possible combinations of the estimated Gaussians. For K=2, we thus also test use of each Gaussian without the other. For K=3, we test use of each Gaussian without the others as well any pair of Gaussians without the third.

The multivariate distributions are evaluated for their potential to produce a statistically viable split by considering the separation of the means using the complement of the formula presented above for merging of adjacent object fragments. That is, the means of two Gaussian distributions are required to be further apart than twice the sum of their standard deviations. When this condition is met, a Bayesian error minimization scheme is used to determine the separation threshold (Duda et al, 2001). The object is then split on the basis of the density values without consideration to their spatial location or any shape related information.

An object may contain millions of voxels. Carrying out the expectation-maximization could be prohibitively costly in those cases. Instead, we generate and use the larger of 10,000 and 10 $\sqrt{N}$ random density values where N denotes the number of voxels in the object. The random number generator is derived from a density histogram of the object.

Figure 7 shows the result of applying Gaussian mixture model based splitting to three different cases. Clean separation is produced in all three cases. Generally speaking, the algorithm performs well when then the object materials have sufficiently different density values and are of comparable size as is the case here. When the object materials have similar density values, the mean separation test will fail. When the objects are of greatly different size, the Bayesian error minimization scheme will favor the larger object at the expense of the smaller object in that the separation threshold will be moved proportionally closer to the mean of the latter, which results in fewer voxels being assigned to it. We speculate that use of voxel location combined with shape related information may allow use of a tighter mean separation test and possibly even counter the tendency to favor larger objects.

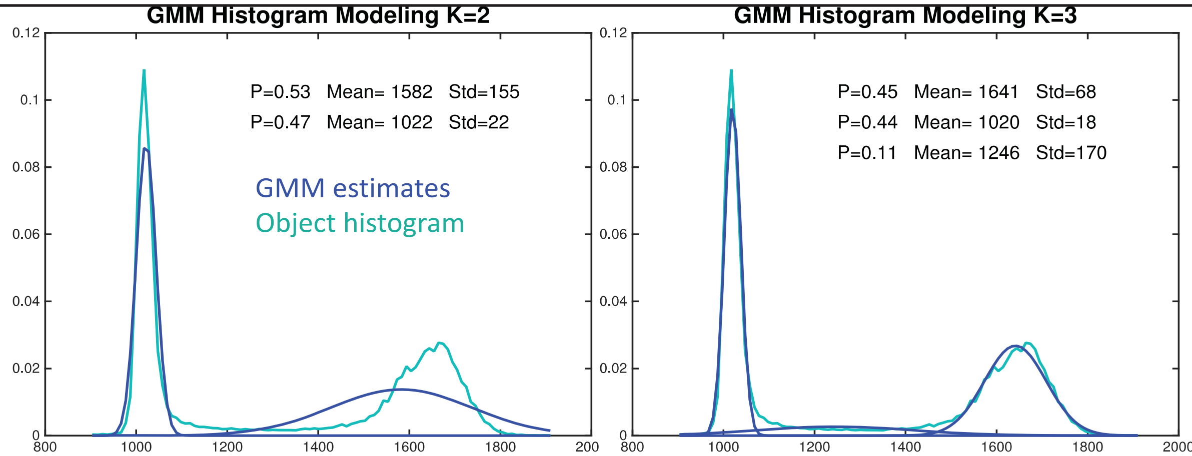


Figure 6. Gaussian mixture modeling of density histograms.

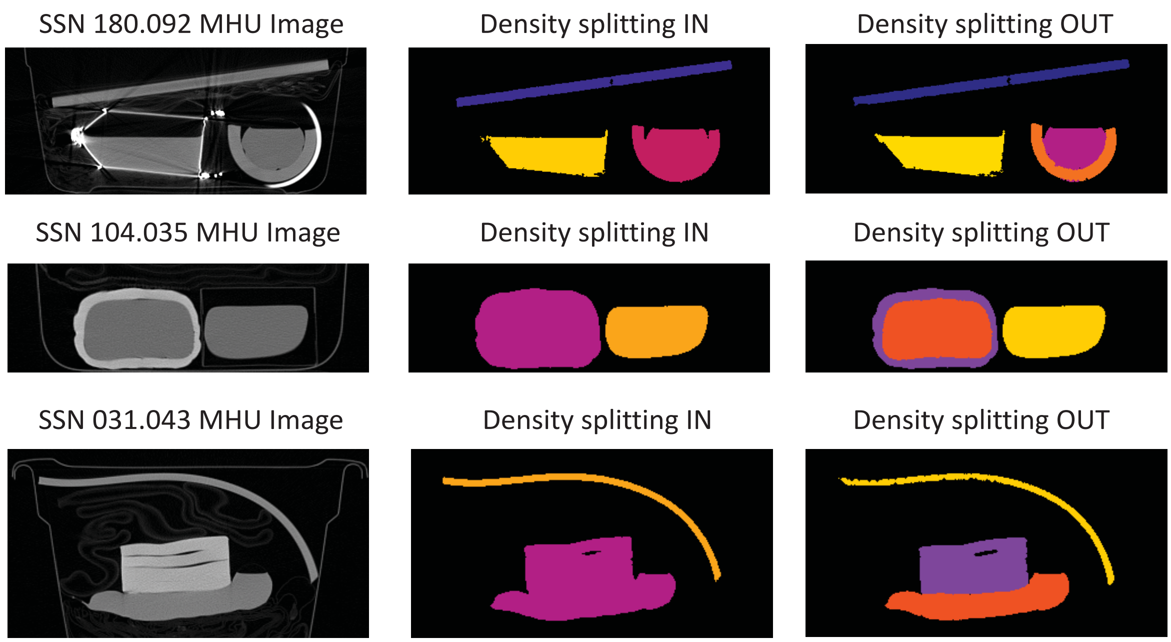


Figure 7. Gaussian mixture model based splitting.

4 Support Vector Machine Based Classification

4.1 Feature Description

The PD/PFA results shown in Tables 1 and 2 were obtained using the following set of features: object size (number of voxels), object mass (weight), object density mean, and the ratio of core voxels located deeper than a set distance to the object size. The first two features are used to eliminate small objects. The first three features are related by the second being a scaled product of the other two. The last feature indicates whether an object is a sheet (value is in the range of 0-0.05) or bulk (value is in the range of 0.05-1.00).

Other features were tested including a wide variety derived from the density histogram. None of these were found to substantially improve the PD/PFA results.

4.2 Support Vector Machine

A support vector machine (SVM) implements a maximum margin classifier in that the separating hyperplane is oriented in such a way that the closest samples from the two classes are as far away from it as possible. In the simplest form, separation is guaranteed of linearly separable feature vectors for a two-class problem (Cortes and Vapnik, 1995). The work presented here used a Gaussian radial basis function to map the data to a non-linear space. Regularization was introduced to allow for imperfect separation (Gregor and Liu, 2004). The resulting constrained quadratic programming problem was solved using the Lagrange dual problem formulation of the primal problem given by

$$\operatorname{argmin} \frac{1}{2}\|w\|^2 + C\sum s_i \quad \text{s.t.} \quad y_i\,(w^T\Phi(x_i) - b) \geq 1 - s_i;\; s_i \geq 0$$

While all other code was written in C++, the fitcsvm function from Matlab was used to handle all aspects of SVM training.

4.3 Classifier Training

Sheet and bulk objects were identified and separated using the aforementioned fourth feature. Training and testing was done independently for these two types of data. Each object was compared against all relevant known targets. The geometric mean of the associated precision and recall values was used to label each training object as either target or non-target. This compensated for such labels not being available for the automatically segmented data.

The purpose of SVM training is to determine separation between the two classes considered. To reduce the potential for overtraining and to make it more likely that class separation favors the targets, kNN based editing was applied to the non-target data. That is, all kNN samples had to be non-targets in order for a non-target sample to be considered during SVM training. The value of k was set to be 3 for bulk objects and large sheets and 5 for small sheets. Figure 8 provides an illustration of this idea. The light blue data represents samples eliminated by kNN editing. The curved nature of the decision boundaries is a result of the Gaussian kernel, which transforms the feature distance comparison used by a linear classifier into a feature similarity comparison that facilitates non-linear separation.

Each feature was subjected to standardization (scaling by standard deviations) prior to training. This was done in order to make distance computations that involve features that have different value ranges meaningful. The cost of misclassifying a target as a non-target was set to be twice that of misclassifying a non-target as a target. This was done to potentially increase PD while accepting the risk of also increasing PFA.

Following standard practice, 10-fold cross-validation was used to ensure separation of training and test data while allowing all data to be classified. The objects were randomly divided into 10 sets. One set was held back for testing while classifier training was carried out using the other nine sets. This process was repeated till all sets had been used for testing.

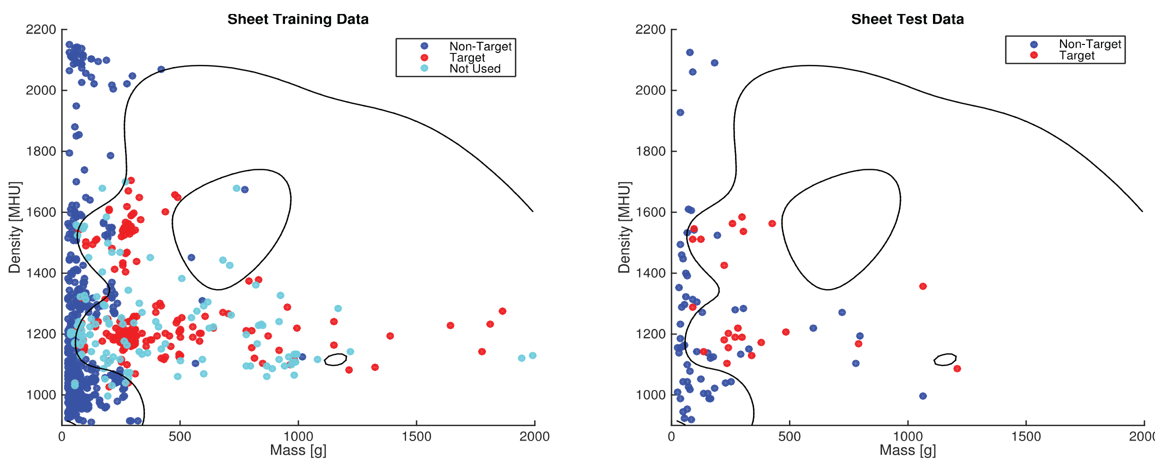


Figure 8. SVM classifier example illustrating kNN editing.

5 Case Studies

Figure 9 shows processing of target 6012, a 10% saline solution in a breast milk bag with a mass of 285 g. The target was detected with a recall of 78% and a precision of 95%. The mass was 205 g. The shading streak artifact is clearly visible but did not affect detection. The object circled in white represents a missed target. This object touched the similarly colored sheet in a difference slice. Both objects were made of rubber.

Figure 10 shows processing of target 6051, a clay object with a mass of 286 g. The target was detected with a recall of 81% and a precision of 89%. The mass was 345g. The object circled in white is the same as mentioned above.

Figure 11 shows processing of target 6150, a clay object with a mass of 290 g. The target was detected with a recall of 70% and a precision of 80%. The mass was 304 g.

Figure 12 shows processing of target 6193, a clay object with a mass of 410 g. The target was detected with a recall of 69% and a precision of 99%. The mass was 298 g. The object circled in white represents a missed detection, which resulted in a false alarm. Detection was missed due to low precision.

Figure 13 shows processing of target 6163, a 5% saline solution in a tin bottle with a mass of 274 g. The target was detected with a recall of 81% and a precision of 89%. The mass was 249 g.

Figure 14 shows processing of target 6018, a 1/4" rubber sheet with a mass of 685 g. The target was detected with a recall of 41% and a precision of 55%. The mass was 603 g. The objected circled in white is the same as the one seen in Figures 9 and 10.

Figure 15 shows processing of target 6144, a 3/8" rubber sheet with a mass of 345 g. The target was detected with a recall of 27% and a precision of 96%. The mass was 95 g. The objects circled in white represent three false alarms. The red arrow points to a missed detection (the clay object is missing from the figure). A different result is shown below.

Figure 16 shows processing of target 7008, a pair of merged rubber sheets with a combined mass of 1360 g. The target was detected with a recall of 90% and a precision of 80%. The mass was 1862 g. The objects circled in white represent two false alarms.

Figure 17 shows processing of target 8026, an 1/8" neoprene rubber sheet with a mass of 350 g. The pseudo-target was missed. The objects circled in white represent two false alarms and a missed detection. A different result is shown below.

Figure 18 shows processing of pseudo target 6026, a powder object with a mass of 277 g. The target was detected with a recall of 97% and a precision of 98%. The mass was 206 g.

Figure 19 shows alternative results for targets 6144 and 8026 using a modified classifier.

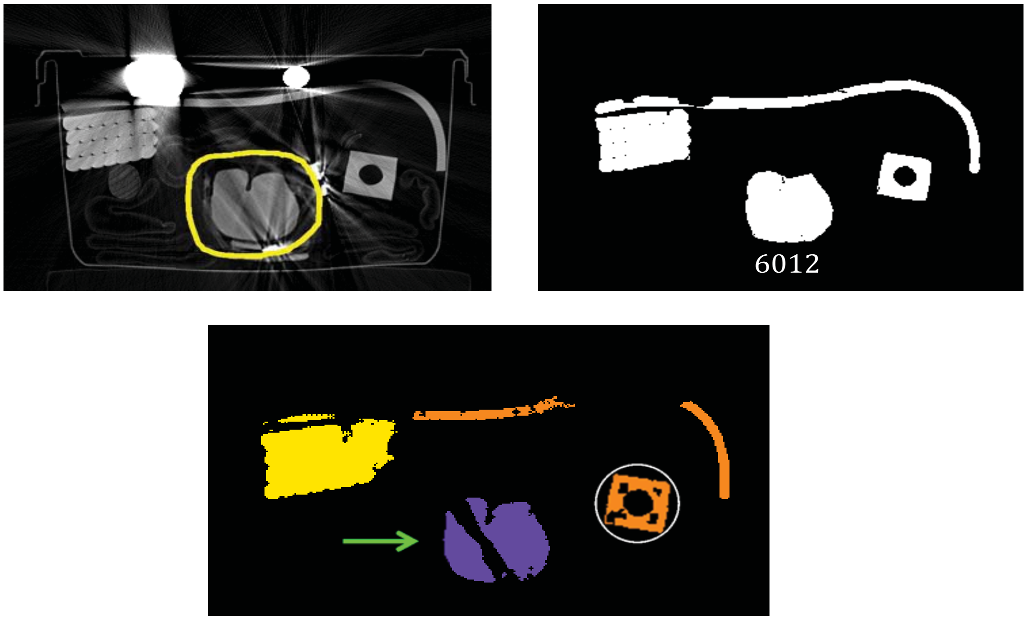


Figure 9. Case 1: SSN 013, slice 105. Target detected.

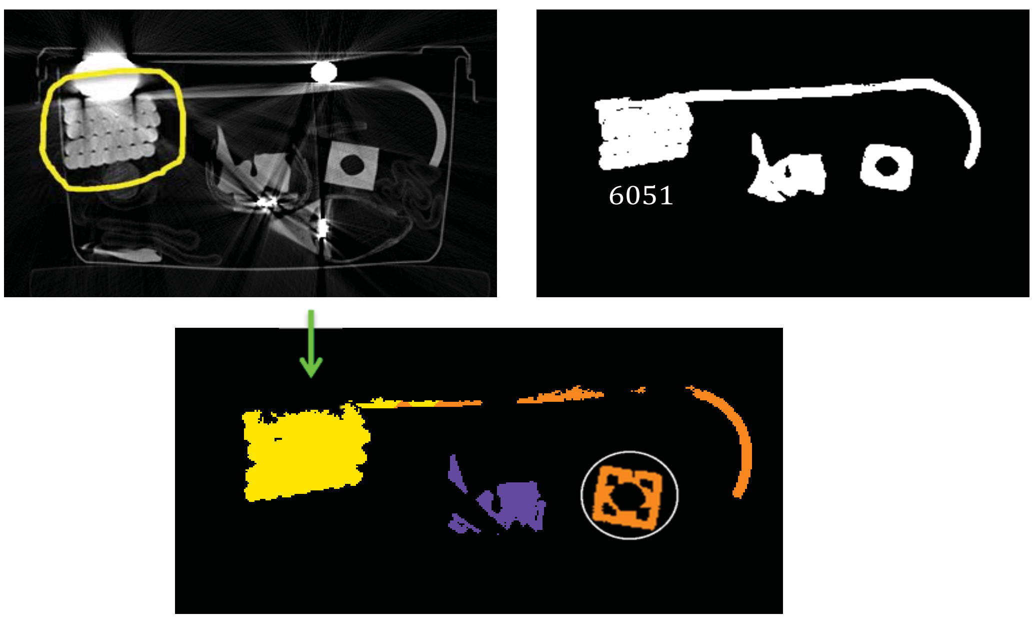


Figure 10. Case 2: SSN 013, slice 128. Target detected.

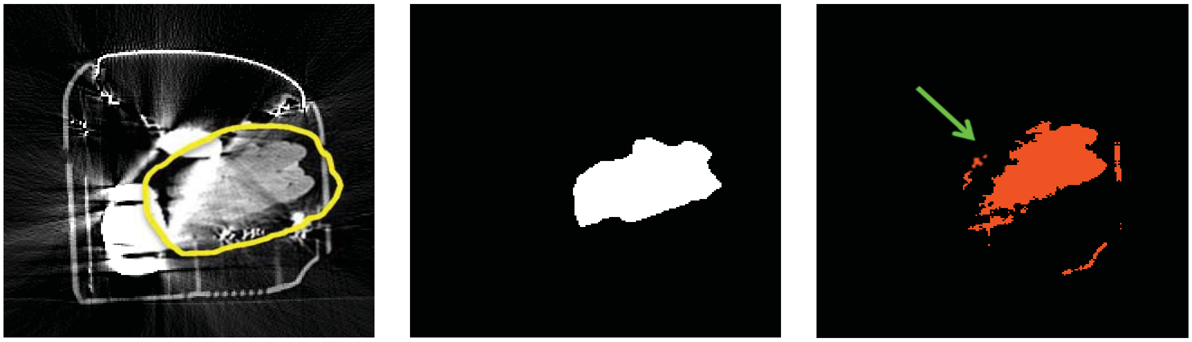


Figure 11. Case 3: SSN 035, slice 049. Target detected.

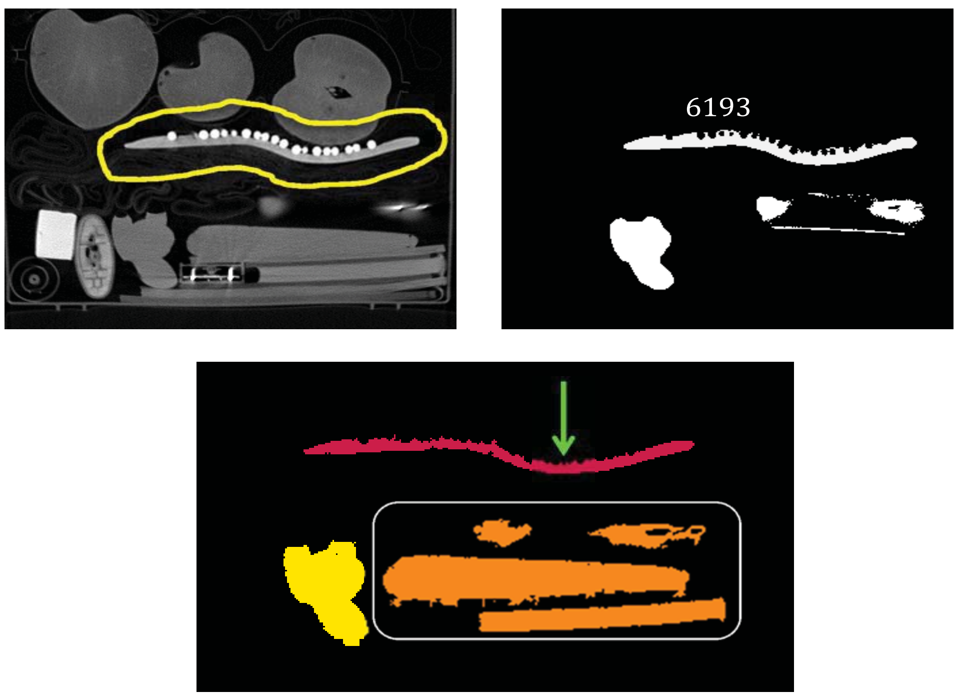


Figure 12. Case 4: SSN 193, slice 198. Target detected.

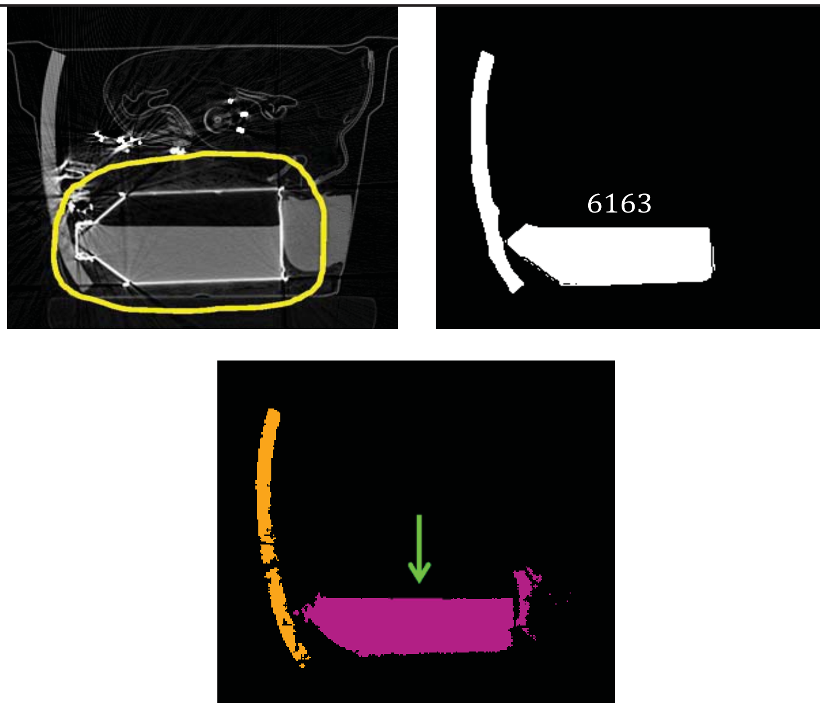


Figure 13. Case 5: SSN 063, slice 045. Target detected.

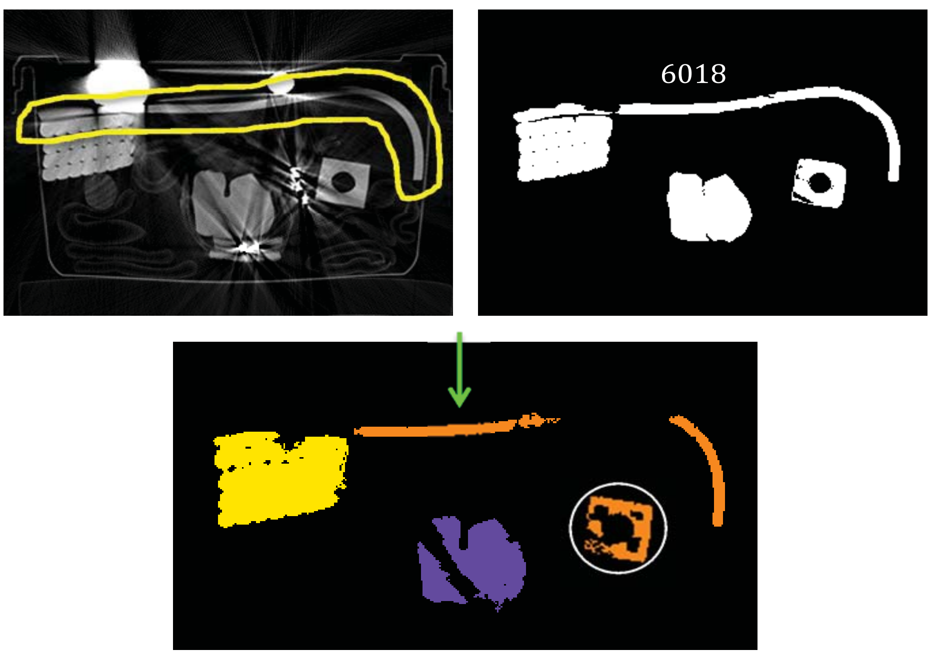


Figure 14. Case 6: SSN 013, slice 111. Target detected.

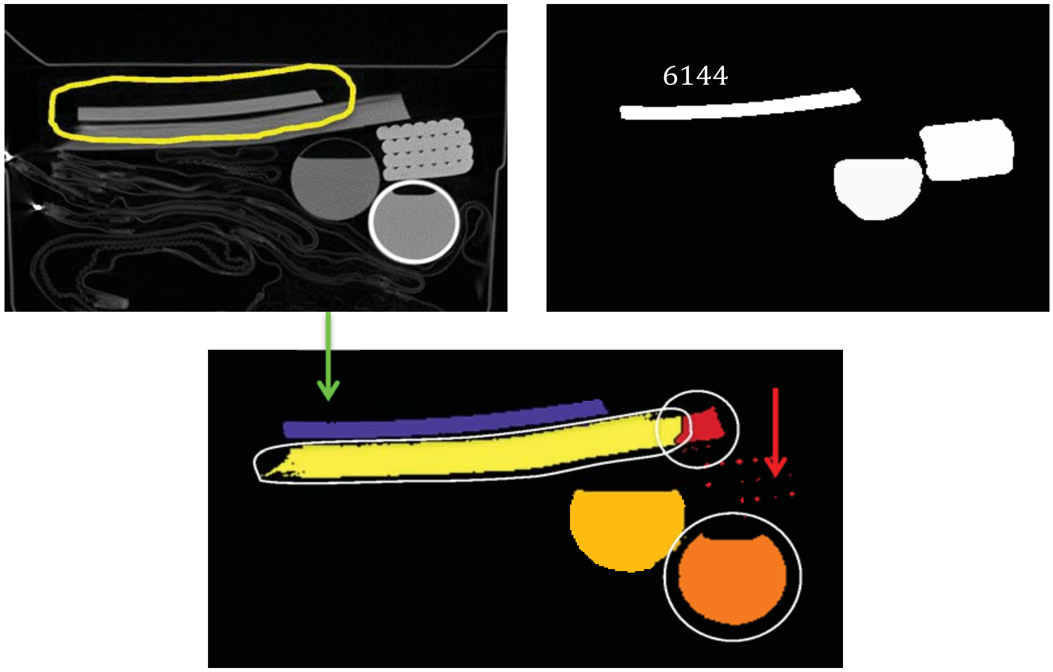


Figure 15. Case 7: SSN 033, slice 046. Target detected.

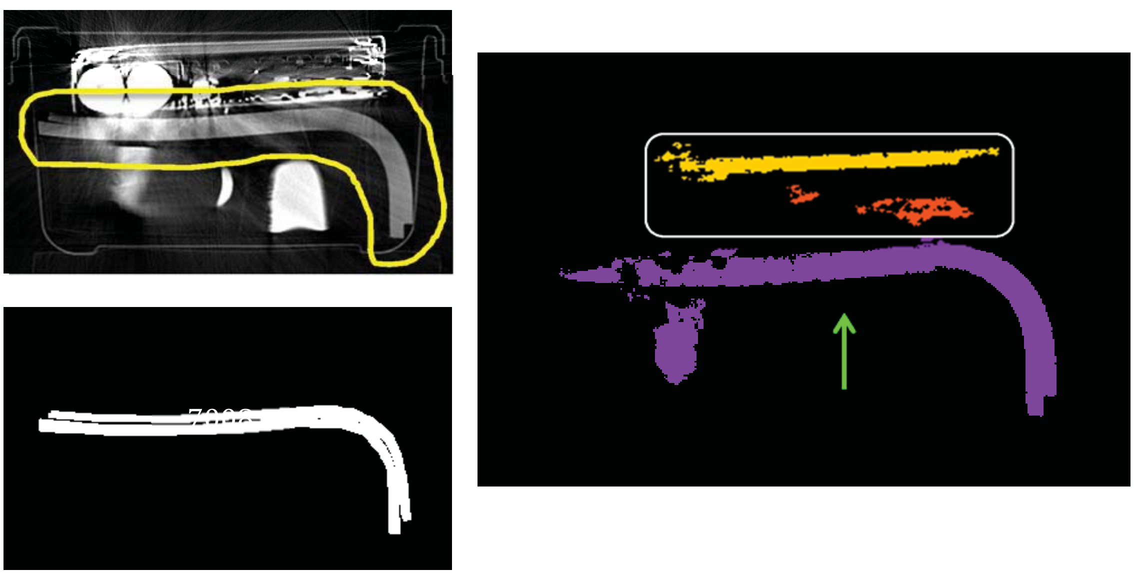


Figure 16. Case 8: SSN 011, slice 094. Target detected.

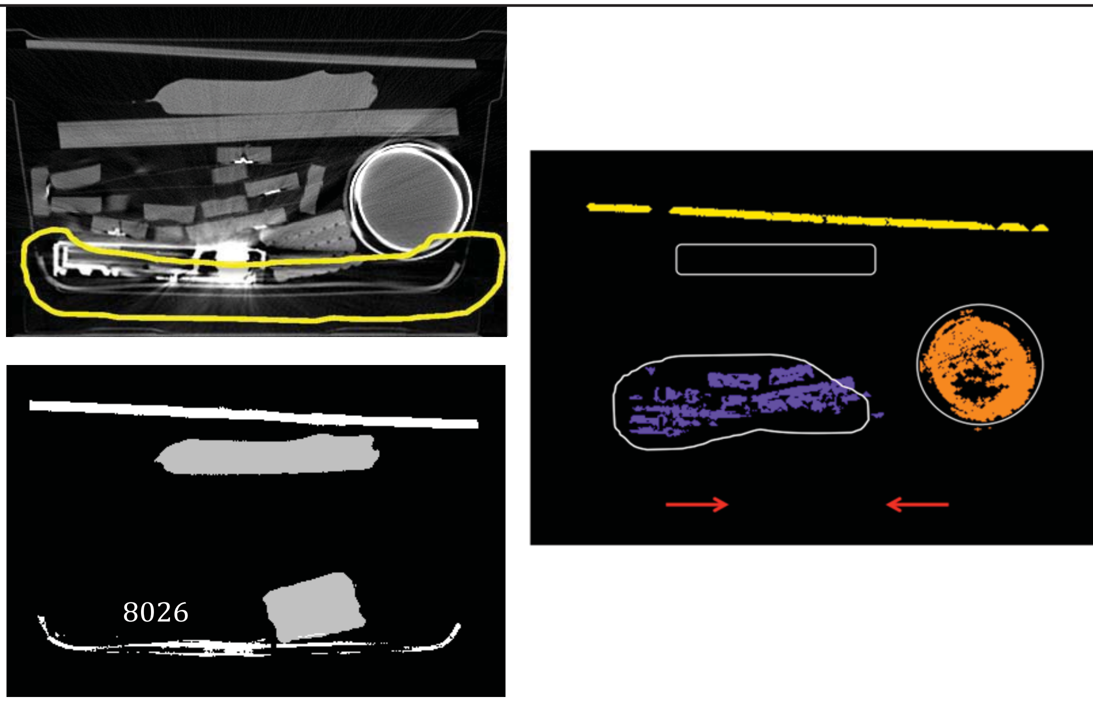


Figure 17. Case 9: SSN 018, slice 125. Target missed.

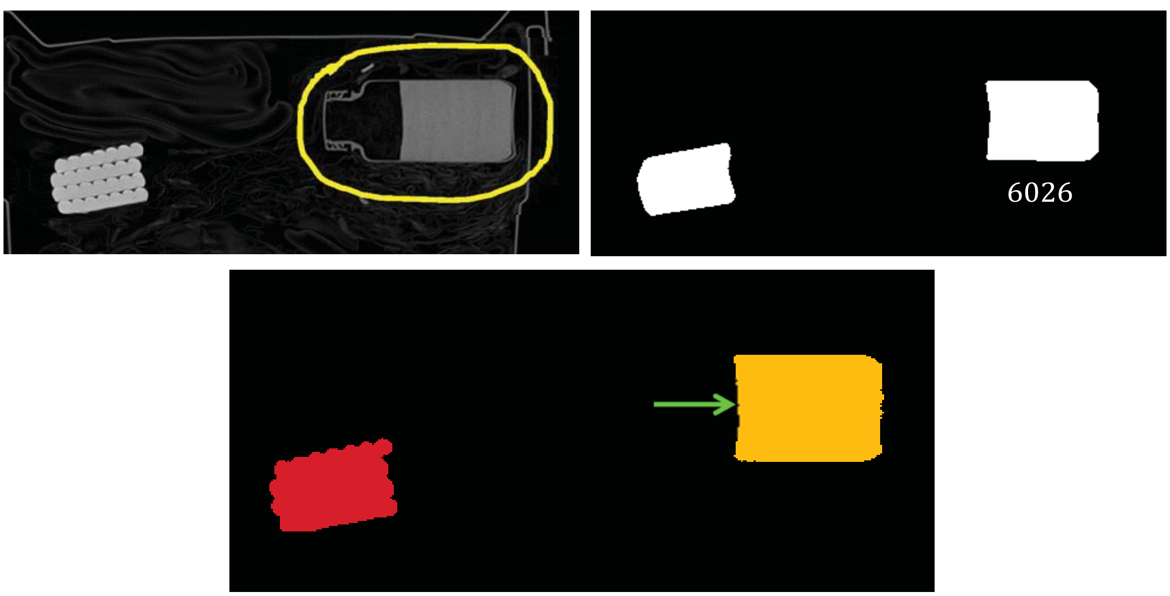


Figure 18. Case 10: SSN 012, slice 115. Target detected.

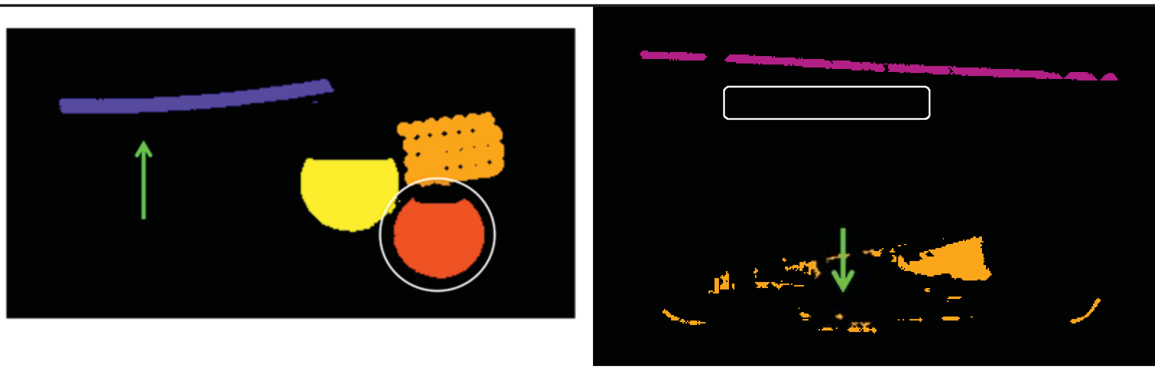


Figure 19. Alternative results for Case 7 (left) and Case 9 (right).

6 Discussion

Numerically, PD is close to the desired 90% while PFA is on the high side at 20%. Figures 9—18 provide object segmentation and labeling illustrations for ten case studies. All but one of the targets of interest were detected. By using more density based features and loosening the convergence constraints on the SVM training algorithm, the PFA was reduced to 15%. However, the PD dropped to 81% overall. As shown in Figure 19, the result for Case 7 was greatly improved in that the clay target previously missed was now included and two false alarms were removed. When furthermore applying the geometric mean of the precision and recall value pairs to weight the training samples, the PD for pseudo-target sheets reached 100%. Figure 19 shows the corresponding detection of the previously missed sheet for Case 9. However, the overall PD dropped to 86% and the PFA increased to 22%. The improvements were thus localized to individual test cases more so than the overall performance.

In order to determine if the high PFA was due mainly to segmentation or classification issues, an experiment was carried out in which the geometric mean of the recall and precision value pair was used as a feature and not just a weight parameter. While this would not be possible in an actual ATR since it depends on the ground truth be known not just for training but also for testing, it does provide valuable insight. The result was an overall PD of 91%, sheet and pseudo-target sheet PDs of 99% and 90%, respectively, and a PFA of 4%. We can thus conclude that the segmentation produces objects that largely meet the given recall and precision requirements. We can also conclude that the automated separation of the training samples into targets and non-targets largely works as intended. This leaves us speculating that the high PFA is mainly a result of the classifier not being able to separate the two classes. We conjecture that more density features are needed to reduce the PFA and that such a move requires an improvement in the quality and consistency of the density distributions. Indeed, analysis of the present data reveals a substantial degree of variation within each target group of objects. For example, the density distributions vary between being unimodal and multi-modal as well as symmetric and asymmetric. Such variation affects the higher order statistics associated with the distributions. One possible course of action might be to have the graph algorithm for shape based splitting take density information into consideration and likewise make the Gaussian mixture model based density splitting algorithm shape aware. Currently, the two operate without knowledge of the other. The goal would be to improve the segmentation accuracy beyond what is needed to meet the given recall and precision requirements. Another possible course of action would be to improve the overall image quality, e.g. by means of metal artifact reduction which was not applied here. This would have a direct impact on the shape and the extent of many of the density distributions.

The advantage associated with the proposed approach was intended to be that accurate segmentation would allow for a simple classifier, which in turn would help prevent overtraining. This is still viewed as being the case. Other improvements to consider for the shape and density based splitting involve relaxation of the rules that govern when and how the decision to split is made. For example, the separation of sheets from bulk objects views sheets as objects whose thickness is between 1/4” and 3/8” while bulk objects are assumed to have core depths greater than 3/4”. By combining the graph based splitting with knowledge of density distributions, it may be possible widen these thickness ranges. Conversely, the Gaussian mixture models do a superb job of indicating when the underlying data is multivariate, but the decision to split currently requires a greater degree of separation between the different mean values than strictly speaking is necessary if the spatial distribution of the data were taken into account as well.

The classifier uses object size and mass as features. Combined with the enclosing nature of the decision boundary produced by an SVM classifier based on Gaussian radial basis function kernel mapping, the end result is upper bounds on size and mass, which is undesirable even if the values are high. A better approach might be to use a greater-than (polynomial decision boundary) classifier for object size and mass to eliminate small objects and then apply a similar-to (enclosing decision boundary) classifier to density related features that capture relevant material characteristics.

Biosketch

Dr. Gregor is a professor of computer science at the University of Tennessee where he has worked on a wide variety of imaging projects over the last 20+ years. He has experience with image reconstruction for modalities ranging from X-ray CT, SPECT and PET to neutron CT. He has worked on applications in the fields of biology, radiology, nuclear medicine, waste management, material science, and most recently security. His academic record consists of more than 70 journal and conference papers on subjects related to imaging and pattern recognition. He is experienced consulting for industry and has written product oriented software. He has taught undergraduate courses in computer organization, C/C++ programming, data structures and algorithms, and software engineering, as well as graduate courses in pattern recognition, image processing and reconstruction, and parallel and distributed computing. Dr. Gregor was a participant in ALERT Task Order 3, which dealt with iterative reconstruction of images of luggage, and ALERT Task Order 4, which dealt with the development of automated target recognition for screening of luggage. He is a US citizen and resides in Knoxville, TN.

Acknowledgements

This material is based upon work supported by the U.S. Department of Homeland Security, Science and Technology Directorate, under Task Order Number HSHQDC-12-J-00056. The views and conclusions contained in this document are those of the authors and should not be interpreted as necessarily representing the official policies, either expressed or implied, of the U.S. Department of Homeland Security.

References

J. Kwon, J. Lee, and S. Song, Extraction of objects from CT images by sequential segmentation and carving, US Patent Application Publication 2013/0170723, July 4, 2013.

C. Cortes and V. Vapnik, Support vector networks, Machine Learning, vol. 20, pp. 273–297, 1995.

R. Gonzalez and R. Woods, Digital Image Processing 3/e, Pearson Prentice Hall, 2008.

T. Cormen, C. Leiserson, and R. Rivest, Introduction to Algorithms, The MIT Press, 1986.

R. Duda, P. Hart, and D. Stork, Pattern Classification, Wiley-Interscience, 2001.

W. Press, S. Teukolsky, W. Vetterling, B. Flannery, Numerical Recipes 3/e, Cambridge University Press, 2007.

J. Gregor and Z. Liu, Regularized SVM parameter estimation, IEEE Intl. Conf. Data Mining, (Brighton, England), pp. 99-105, 2004.

11.6.3 ATR Development – Zhang

"MRF-EM Image Segmentation and Collective FA Classification for Target Detection in CT Scanned Bag Images – University of Wisconsin-Milwaukee Final Report for TO4: Automatic Target Recognition Initiative"

MRF-EM Image Segmentation and Collective FA Classification for Target Detection in CT Scanned Bag Images – University of Wisconsin-Milwaukee Final Report for TO4: Automatic Target Recognition Initiative

Jun Zhang, Laura Drake, and Hongquan Zuo
Department of Electrical Engineering and Computer Science
University of Wisconsin-Milwaukee
Milwaukee, WI 53211
Contact:junzhang@uwm.edu, 414-229-4246

Abstract

In this report, we describe an ATR system for detecting targets in 3D CT scan images of luggage bags. The main types of targets to be detected are: sheet rubber, bulk rubber, saline, and clay. Our ATR systems consists of two parts, image segmentation and false alarm detection. In image segmentation, pixels in an input bag image are grouped into a set of 3D regions of relatively homogenous intensity levels, potentially corresponding to targets/objects in the image. In false alarm detection, regions generated from segmentation are further examined and those that correspond to non-targets are detected and removed. Our image segmentation algorithm is based on an MRF-EM (Markov random fields and expectation-maximization) pixel classification technique while our false alarm detection algorithm is based on a collective classification technique using SVMs (support vector machines); both are novel with respect to prior work as known to us (e.g., prior patents provided by the management). On the project test image data set, our ATR system achieved the management goal of 90% PD (probability of detection) and 10% PFA (probability of false alarm), with a PD of 80% on PT (pseudo target) sheets.

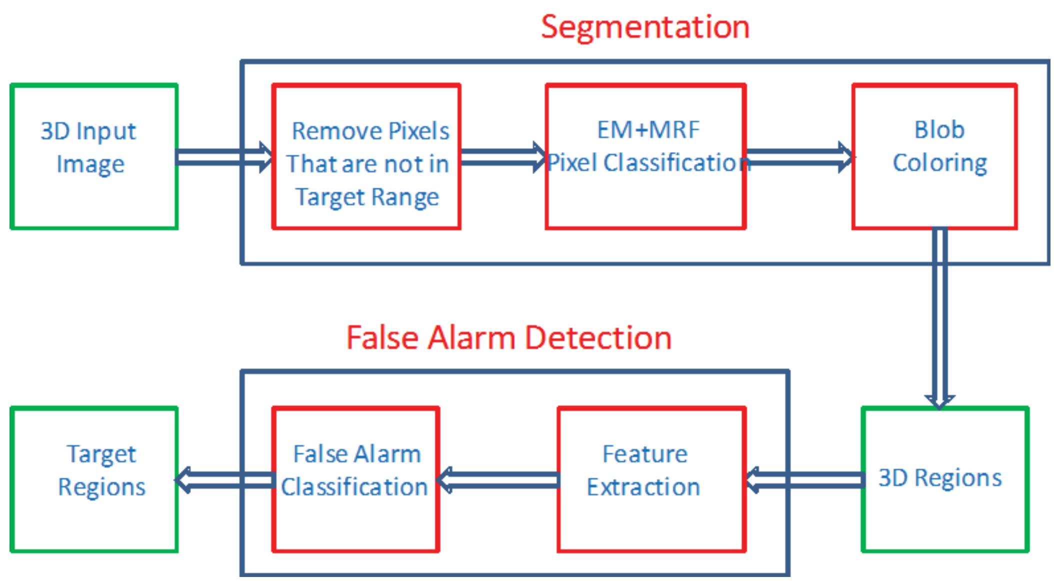


Figure 1: Our ATR System: an Overview.

1. Introduction

According to some statistics, each day in the United States there are 2 million passengers traveling by air. Assuming half of them have one bag or suitcase to check, the average number of checked bags each day could easily reach a million. Currently, these bags may be scanned by CT scanners and the resulting 3D images are processed by automatic target detection (ATR) algorithms to check for suspicious objects or materials, especially explosives. If such targets are detected in the image of a bag, the bag may need to be manually examined and processed. Today's state-of-the-art commercial algorithms and systems, which passed government certification tests, generally do a good job in ATR, with acceptable detection and false alarm rates. However, according to a recent report [1], the Department of Homeland Security (DHS) may require future ATR systems to have increased detection and reduced false alarm rates for a larger set of targets and with reduced minimum masses. Furthermore, there are indications that such future requirements may be difficult to achieve with the technologies employed by current commercial algorithms and systems. The purpose of the research reported here is to develop and evaluate new ATR algorithms that can potentially meet the DHS future requirements. If successful, our results can potentially contribute to the further improvement of air travel security.

Fig. 1 provides an overview of our approach and our ATR system. The ATR system takes a 3D CT scanned bag image as input and produces detected target regions (if any), also known as a label image, as output. If there are target regions in the output, an alarm is associated with the bag. For this project, there are four main target types of interest: clay, bulk rubber, rubber sheets, and saline. In addition, there are also a number of pseudo target types, such as thin rubber sheets.

As shown in Fig. 1, our ATR system consists of two processing blocks, image segmentation and false alarm (FA) detection. In the image segmentation block, pixels in the input CT image are

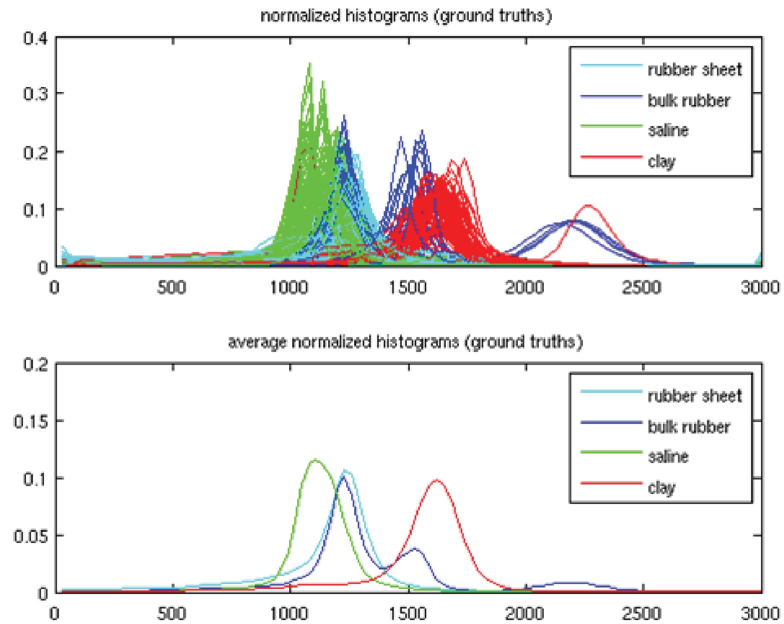


Figure 2: Normalized Histograms of Ground Truth (GT) Targets and their Averages. Light blue: rubber sheet, dark blue: bulk rubber, green: saline, red: clay.

grouped into several 3D regions that hopefully correspond to objects in the bag. In the FA detection block, these regions are classified based on their features to determine if they are false alarms, i.e., non-targets. Regions that are non-targets are removed while regions that are not classified as false alarms are declared as detected targets for the bag. ATR tests on the project image data set indicate that our system met project targets: a particular realization of the system achieves a probability of detection (PD) of 89.2%, with a probability of false alarm (PFA) of 9.7% and a pseudo target (PT) sheet PD of 80%. This achieves the nominal management goal of 90% PD and 10% PFA.

The rest of the report is organized as follows. In Sections 2, 3, and 4, respectively, we describe our image segmentation and FA detection algorithms as well as relevant prior work. In Section 5 and 6, respectively, we present ATR results and provide conclusions and directions for future work.

2. Image Segmentation

2.1. Overview

Image segmentation takes a CT scanned bag image as input and generates, as output, a set of 3D regions that hopefully correspond to targets or other objects in the bag. To develop our segmentation algorithm, we started by examining the project image data set, which contains 188 3D CT bag images, and our observations have been quite useful. For example, Fig. 2 shows the (normalized) pixel intensity histograms of the ground truth targets (which were obtained manually [i.e., manually

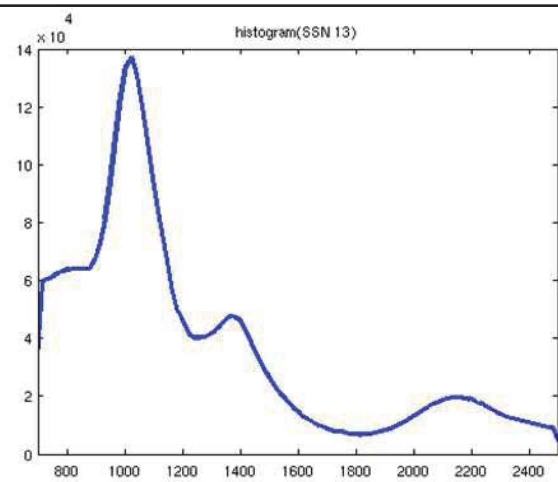


Figure 3: Histogram of an Image (SSN 13).

segmented and identified] from the images by Franco Rupicich). From these we can make the following observations:

1) Most target pixels have intensities in the range of $[a, b]$, e.g., with $a = 700, b = 2500$.

2) The target pixel histograms are mostly Gaussian-like with a single peak (except for bulk rubber whose histogram looks like a mixture of two Gaussians, see Fig. 2, bottom figure, dark blue curve)

3) Some target histograms overlap (e.g., for saline and rubber, see Fig. 2, bottom figure, the overlap between the green and blue curves)

4) Imaging artifacts tend to increase "the spread" (or variance) of the target histograms

Similarly, Fig. 3 shows the histogram of a typical image in the image set (SSN 13). From this and the histogram of other images we can make the following additional observations:

5) The histogram of the image contains multiple Gaussian-like "peaks" or "hills"

6) Each peak is generally made up by pixels from a single object (target or non-target) or, multiple objects with similar intensity histograms (i.e., with similar means and variances)

Based on these observations, we developed a 3-step segmentation algorithm, as shown in Fig. 1. Since most target pixels have intensities in $[700, 2500]$ (see Fig. 2), the Step 1 of our algorithm labels pixels whose intensities are outside this range as "background" and remove them from further consideration. Then, in the Step 2 of our algorithm, the remaining pixels (i.e., foreground pixels) are classified according to their intensities and certain spatial constraints (detailed later). Finally, in Step 3, spatially connected pixels in the same class are grouped together into regions, using the "blob coloring" algorithm [2] (also known as the connected component linking [CCL]). Since Step 1 and 3 are relatively straightforward, we will focus our description on Step 2: pixel classification.

First, we provide an intuitive description.

2.2. Pixel Classification: Intuitive Description

From our observations in Section 2.1 and Figs. 2 and 3, the histogram of an image is made up by several and possibly overlapping Gaussian-like peaks; each peak generally correspond to a single object or multiple objects with similar intensity distributions. Hence, one way to segment the image and form/detect objects is to associate each pixel of the image with one of the peaks of the histogram; in other words, each pixel is classified as to which histogram peak it belongs to. In this way, spatially connected pixels that belong to the same peak, or *class*, can be grouped into potential object regions.

Since the histogram peaks have Gaussian-like shapes, they can be described/modeled by Gaussian distributions, each with a mean and a variance. Hence, in practice, the pixel classification process is a two-step process: first, estimate the Gaussian distribution parameters and then, classify each pixel using the estimated Gaussian parameters/distributions. To boost results, this two-step process is usually iterated a number of times: Gaussian distribution parameter estimation, pixel classification, Gaussian parameter estimation again, pixel classification again, so on and so forth, till some type of convergence is reached (e.g., the classification results no longer change significantly). This is the basic idea behind the EM (expectation-maximization) algorithm.

Another aspect of the pixel classification for segmentation is that due to noise, artifacts, and overlap between the histogram peaks, if each pixel is classified independently, i.e., using only its own intensity information, the classification results tend to be "noisy" – object regions will have a lot of small holes and breaks. This problem can be solved by imposing some spatial "smoothness-constraints" in the iterative two-step classification process described above and one way to do this is to use an Markov random field (MRF). Intuitively, the MRF encourages neighboring pixels to be classified into the same class, thereby removing small holes and breaks.

Next, we provide a more precise description of our pixel classification algorithm.

2.3. Pixel Classification: Mathematical Description

First, we introduce some notation. Let $y = \{y_i\}$ be a 3D input CT image, where y_i is the pixel (voxel) at location i. Suppose the image's histogram contains K "peaks", or *classes*. Let $z = \{z_i\}$ be a set of class labels assigned to image y, where z_i is the class label of y_i and $z_i = k$ indicates that y_i is associated with class k.

Since our images contain random objects (e.g., in addition to targets, there are books, water bottles, and other random objects) and since the imaging process is full of random factors, such as imaging noise and metal artifacts, we model both y and z as random variables, with joint probability distribution

$$p(y, z) = p(y|z)p(z) \tag{1}$$

Here, $p(y|z)$ is a conditional distribution and we assumed

$$p(y|z) = \prod p(y_i|z_i) \tag{2}$$

Intuitively, $p(y_i|z_i)$ models the kth histogram peak of the image when $z_i = k$ and, since the histogram peaks are modeled by Gaussian distributions, $p(y_i|z_i)$ is a Gaussian distribution. Similarly, for $p(z)$, we assume for the moment an independent distribution with

$$p(z) = \prod p(z_i) \tag{3}$$

with

$$p(z_i) = \pi_k \quad \text{if } z_i = k \tag{4}$$

where

$$\pi_k > 0 \quad \text{and} \quad \sum \pi_k = 1 \tag{5}$$

are a set of (probability) parameters.

Intuitively, this simple model suggests that the classification of each pixel is independent from that of other pixels. Finally, we use parameter vector θ to contain all the parameters associated with $p(y|z)$ and $p(z)$, including the means and variances of the Gaussians (for $p(y_i|z_i)$) and π_ks (for $p(z_i)$).

Using the above notation and models, the pixel classification problem can be described as: finding θ and z from y. A good solution to this problem is the EM algorithm [3], which seeks to solve the following maximum-likelihood problem:

$$\hat{\theta} = \arg\max_{\theta} \log p(y|\theta) = \arg\max_{\theta} \log \left\{ \sum_{z} p(y, z|\theta) \right\} \tag{6}$$

The EM algorithm iterates between the following two steps:

The E-step: compute the class-probability of each pixel

$$r_{i,k} = P[z_i = k] = Cp(y_i|z_i = k, m_k, \sigma_k)\pi_k \tag{7}$$

where m_k and σ_k are the mean and variance of the kth class (kth histogram peak), π_k is the proportion of the kth class [see also eqns (3) and (4)], and C is a normalization constant

The M-step: compute the mean, variance, and proportion for each class (i.e., each histogram peak)

$$m_k = \frac{\sum r_{i,k} y_i}{\sum r_{i,k}}, \quad \sigma_k^2 = \frac{\sum r_{i,k}(y_i - m_k)^2}{\sum r_{i,k}}, \quad \pi_k = \frac{\sum r_{i,k}}{n} \tag{8}$$

where the sum is over all i, i.e., over all pixels.

When the EM algorithm converges, we can obtain the final pixel classification by thresholding the class-probability of each pixel, with

$$z_i = \arg\max_{k} \{r_{i,k}\} \tag{9}$$

Intuitively, in the EM algorithm, the E-step estimates a pixel's probability of belonging to each of the K histogram peaks or classes of the image. Compared to assigning the pixel exclusively to a single class, this "soft-decision" takes into account that the intensity range of the different classes can overlap, and this leads to better final classifications. Similarly, the M-step updates the Gaussian

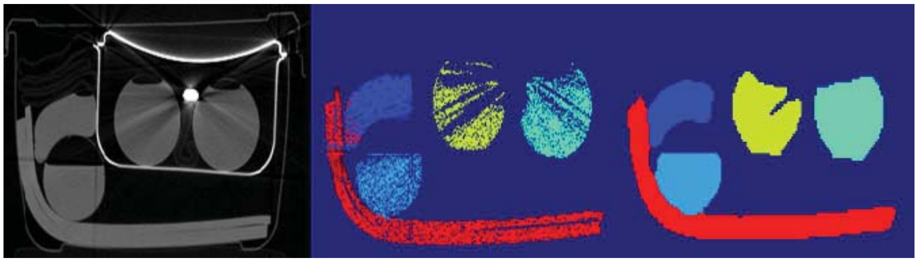


Figure 4: Some Image Segmentation Results in Slices. Left: input image, middle: EM-only segmentation (noisy), right: MRF-EM segmentation (smooth).

parameters for the histogram peaks and iterations between the E and M-steps boost classification results, as described in Section 2.2. Finally, the EM algorithm needs an initialization, i.e., initial parameters for θ. This can be obtained by running a common clustering algorithm, such as the K-means [5], on the image data first [7].

2.4. Pixel Classification: Adding the MRF

Although the EM algorithm is very effective, when applied alone, because of its independent assumption on z [eqn (3)] it tended to generate noisy pixel classifications and segmentations with small holes and breaks, see, for example, Fig. 4. A solution to this problem is to introduce spatial smoothness constraints through an MRF [6]. Specifically, this amounts to replacing the independent model of $p(z)$ in eqn (3) by that of an MRF, with

$$p(z) = \frac{1}{Z} e^{-\beta U(z)} \tag{10}$$

where $\beta > 0$ is a control parameter ("the temperature"), Z is a normalization factor, and $U(z)$ is an "energy function," defined as

$$U(z) = \sum V(z_i, z_j) \tag{11}$$

where the sum is over all neighboring pixel location pairs i and j. The function $V(z_i, z_j)$ is known as a clique function and in our case, we use

$$V(z_i, z_j) = \begin{cases} -1, & \text{if } z_i = z_j \\ +1, & \text{if } z_i \neq z_j \end{cases} \tag{12}$$

Intuitively, this function imposes a smoothness constraint: the classifications (z's) where most neighboring pixels belong to the same class have lower energy and higher probability to occur.

Using the MRF $p(z)$ to replace the independent $p(z)$ in eqn (3), the EM algorithm is modified into an MRF-EM algorithm. Based on the derivations in [7], only the E-step is changed (the M-step stays the same), with the new E-step consists of finding

$$r_{i,k} = P[z_i = k] = C \sum_{z \setminus z_i} p(y_i | z_i = k, m_k, \sigma_k) p(z) \tag{13}$$

where C is a normalization constant and the sum is over all possible configurations of z with z_i taken out. In general, this sum is exponentially complex (e.g., for an n pixel image and assume there are only two classes, the number of terms in the sum is 2^{n-1}) and approximations have to be used. An effective approximation technique for this is the mean field theory (MFT) [11]. Originally from statistical mechanics, the MFT has since found a number of applications in image processing in the 1990s to 2000s and currently, it is enjoying a revival in machine learning applications, especially in the so-called variational techniques [12]. In previous work, we derived a set of MFT equations for pixel classification for 2D image segmentation [7]-[10]. For this project, we have extended them for 3D image segmentation. However, since their derivations are relatively lengthy and since they are relatively straightforward extensions of the 2D case, we refer the readers to [7]-[10], where the 2D case is derived for various situations.

To end this section, we provide an example in Fig. 4, where the MRF-EM is used for image classification. It provides a more smooth classification/segmentation and is robust to some of the metal artifacts.

2.5. Segmentation Algorithm Recap and Related Issues

To recap, our segmentation algorithm can be summarized as follows:

Input: a 3D CT image of a bag

Output: 3D regions that generally correspond to objects (including targets) in the bag

Step 1: Remove pixels outside the target intensity range as background

Step 2: Perform MRF-EM classification on the remaining (i.e., foreground) pixels, using eqns (13) and (8)

Step 3: Use blob coloring (i.e., CCL) to group spatially adjacent same-class pixels into regions (the CCL in this case has only one "parameter," the number of 3D neighbors for a pixel; in this case, it is 26, including the pixel's direct and diagonal neighbors)

A typical segmentation result after steps 1 to 3 is shown in Fig. 5, where the resulting regions consist of targets and non-target objects whose intensities are in the target range.

Finally, the project management raised a number of issues related to our segmentation algorithm and these are addressed below:

1) Prevent over-training: our segmentation algorithm is unsupervised and does not require any training. Hence, there is no over-training.

2) Split-and-merge of regions: this is addressed implicitly through the parameters of the MRF-EM classification algorithm. For example, a large K (the number of classes) and small β (for the MRF) encourage large non-homogeneous regions to split while a moderate value of K and a large β encourage small regions to merge. Currently, these parameters are selected through experimentation.

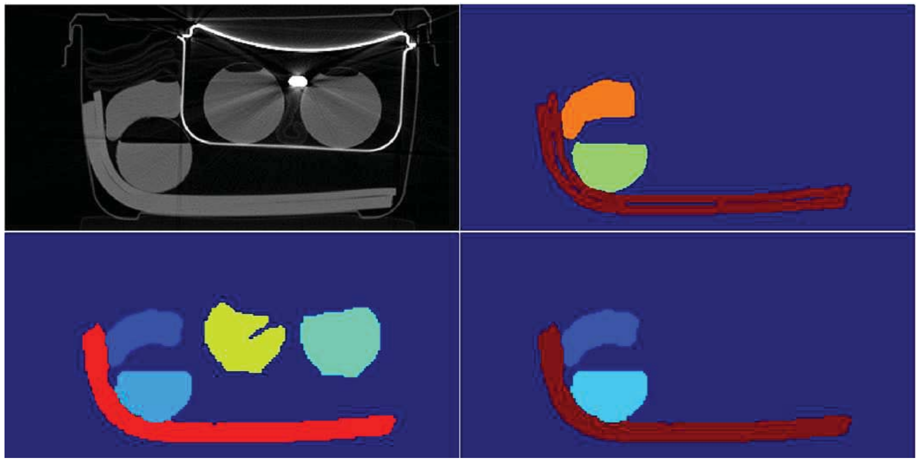


Figure 5: Image Segmentation: a complete result. Top row: original image; ground truth. Bottom row: MRF-EM segmentation + blob coloring (with targets and non-target objects); after FA detection (non-targets removed).

3) Detect and correct artifacts: since the results of our segmentation algorithm is sufficiently good (with PD in the 90% range), this was not incorporated; however, it could be incorporated in the future to further improve our segmentation performance (to achieve PD in the 95% range).

4) Address bulks and sheets differently: not done; for now we want our algorithm to be sufficiently general and not to have special processing units

5) Special processing for pseudo target sheets: for the same reason as 3), no attempt is made in this regard; however, our results are quite good on pseudo target sheet detection (can reach 80% PD)

In summary, we have kept our segmentation algorithm sufficiently general and avoided various special processing units. This approach has two advantages: 1) the algorithm is broadly applicable and is insensitive to changes in the target and non-target objects (as long as their histograms have Gaussian-like shapes) and 2) the main algorithm is a single block (MRF-EM) and it is easier to understand on an image why it works (or does not work).

3. FA detection

Given an input image, our image segmentation algorithm can potentially generate two types of regions, targets and non-targets. The latter generally consists of non-target objects, such as bottled water, and spurious regions, such as small pieces from a laptop computer which has highly varying intensities caused by metal artifacts. If kept, they produce false alarms and lead to high ATR system PFA. Our FA detection algorithm is developed to detect and reject them. As illustrated in Fig. 1, it contains two parts, feature extraction and FA classification.

In feature extraction, we collect/compute features for each region produced by image segmentation. These include region-based histogram, gradient, and other features. For a given region, the histogram features consist of the histogram of the pixels in the region, with 61 bins, and the histogram's mode, mean, and variance. The gradient feature is the norm of the 3D image gradients averaged over the entire region, and the other features include region size, mass, and the ratio of mass to size. In summary, for each region, we extract a total of 68 features (as just described) and based on these features, FA classification determines if a region is "a false alarm" or "not a false alarm." This is a classical classification problem and many effective algorithms, i.e., classifiers, can be used [12].

3.1. Single Classifier False Alarm Detection

In this project, we started with the SVM (support vector machine) as the classifier. Intuitively, the SVM is a hyperplane placed in the input data space (or feature space) that separates the two data classes to be classified. It is very effective and popular in many classification and machine learning applications, thanks to several factors: 1) it is a "max-margin" classifier [12] which makes it robust, i.e., generalizing well on testing data, 2) it can incorporate the "kernel-trick" [12] to separate/classify data classes that are not linearly separable, and 3) its training is a quadratic programming problem that has a unique solution and has fast algorithms. Currently, the SVM has become an off-the-shelf algorithm and after the features are defined, its training is quite simple (e.g., can be done with a few lines of Matlab or R code).

In this work, we randomly took 50% of all the regions (from all images) produced by our segmentation algorithm and used their features to train an SVM. After training, the SVM is used to classify the segmented regions for each image. During training, we can produce a ROC (receiver operating characteristics) curve for the SVM, as illustrated in Fig. 6 (the ROC curve is produced by setting a set of thresholds for the SVM, each will generate a point in the ROC curve). Let (pd, pfa), (PD_{SEG}, PFA_{SEG}), and (PD_{ATR}, PFA_{ATR}) denote, respectively, the PD and PFA for the SVM, the segmentation algorithm (before FA detection), and the ATR system (after FA detection). Then, they are related by

$$PD_{ATR} = PD_{SEG}(1 - pfa), \quad PFA_{ATR} = PFA_{SEG}(1 - pd) \tag{14}$$

This means that we can select an SVM with a particular pd and pfa to achieve an overall ATR system PD and PFA. Generally, we want to tune the segmentation algorithm such that PD_{SEG} is high (e.g., 90% or as high as possible). Then, we want to use an SVM with a small pfa (e.g., 3%) and a large enough pd (e.g., 80%) to remove most false alarm regions without removing too many targets. For example, in one such configuration of our system, image segmentation produced $PD_{SEG} = 89\%, PFA_{SEG} = 62\%$; using a SVM with $pfa = 3\%$ and $pd = 80\%$, we obtained a final system PD of 86% and PFA of 12%.

3.2. Collective Classification for FA Detection

By examining the histograms of all the segmented regions, as shown in Fig. 7, we can take one step further in developing a better FA detection algorithm than the one in Section 3.1. Indeed, Fig. 7

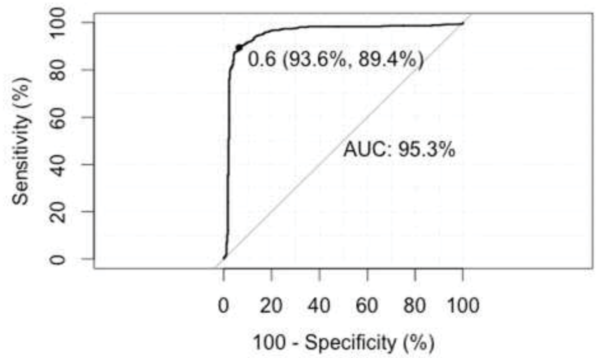


Figure 6: An FA detector's ROC curve. The horizontal and vertical axes are, respectively, are the *pfa pd*. A good operating point: pfa=3%, pd=80%

shows that while the target regions' histograms occupy different intensity intervals, the false alarm regions' histograms are all over the entire target intensity range. This suggests a collective classification algorithm, with multiple classifiers and one for each "target interval." Intuitively, such a collective classification strategy could be seen as an example of "divide and conquer." Indeed, it may be easier to develop/train several classifiers, each with a limited task scope, than to develop/train a single classifier with a very broad task scope.

Now, our collective classification algorithm can be described as the following steps:

The Collective Classification Algorithm

1) Divide the foreground intensity range (e.g., [700, 2500]) into several (possibly overlapping) target intervals, one for each target type. For example, since the histograms of rubber sheet targets are mostly in [915, 1600], this will be used as the target interval for rubber sheets. Similarly, we can obtain the target intervals for the other three target types, as shown in Fig. 8.

2) Associate a classifier, e.g., an SVM, with each target interval. This results in 4 classifiers.

3) For each segmented region, if its histogram mode falls into a particular target interval, send it to the classifier associated with that target interval. Due to target interval overlap, a region 's histogram mode can fall into more than one target interval; in this case, it is sent to more than one classifiers. For example, suppose a region's histogram mode is 1200, it falls into 3 target intervals (rubber sheet, bulk rubber, and saline), and is sent to the 3 classifiers.

4) If a region is sent to a single classifier, it is classified in the normal way, i.e., "false alarm" or "not a false alarm." If it is sent to multiple classifiers, as in the example above, it will only be classified as a false alarm if all classifiers say it is (or, equivalently, it is classified as a target [not a false alarm] if at least one classifier says it is).

The training of the collective classifiers is similar to that for the single classifier, except now the training data is divided into several subsets. Specifically, we

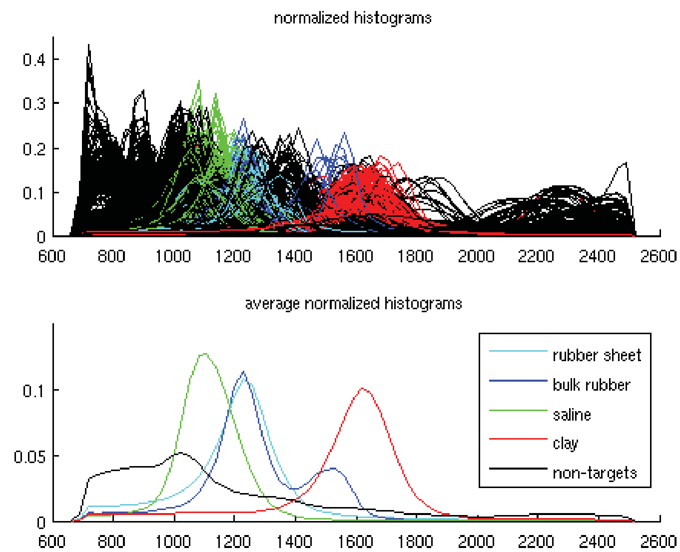


Figure 7: Histograms of Segmented Regions and their Averages. Light blue: rubber sheet, dark blue: bulk rubber, green: saline, red: clay, black: false alarm regions.

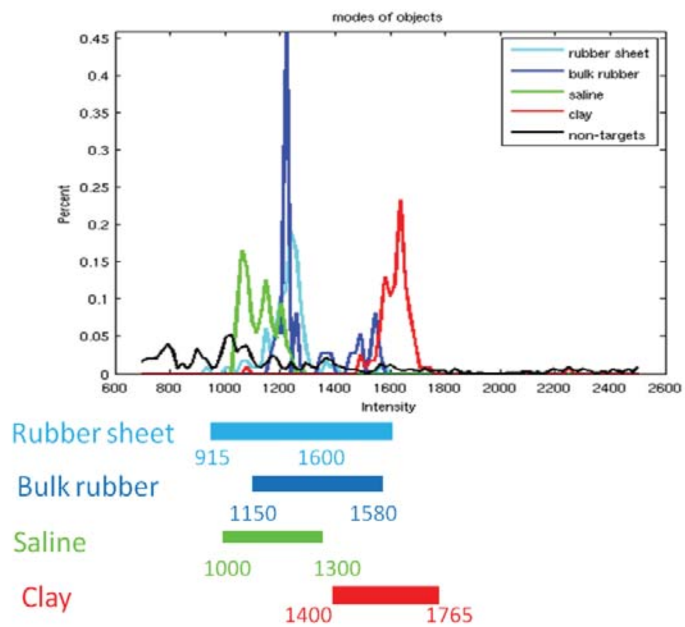


Figure 8: Distributions of the Histogram Modes for the Segmented Regions and Target Intervals. Light blue: rubber sheet, dark blue: bulk rubber, green: saline, red: clay.

12

Collective Classifier Training Algorithm

1) First collect the random training data the same way as described in Section 3.1.

2) Divide the training data into 4 subsets, corresponding to the 4 target intervals. A training region will belong to a particular “target subset” if its histogram mode falls into that target interval. Similarly, a training region can belong to multiple target subsets if its histogram mode falls into multiple target intervals.

3) The kth training subset will be used to train the kth classifier, with $k = 1, 2, 3, 4$.

4) For the classifiers, we used the kernel SVM with 68 inputs (accounting for the 68 region-based features) and the standard Gaussian kernel.

In a typical run of our ATR system, we had a 92% PD and a 42% PFA after segmentation (but before FA detection). After collective classification FA detection, as described above, we had a 90% PD and a 9% PFA.

3.3. FA Detector Recap

In summary, the collective classification algorithms of Section 3.3. is used in our ATR for FA detection. The steps of these algorithms are:

Step 1: Extract region based features (detailed in the beginning of Section 3

Step 2: Training the collective classifiers (algorithm detailed in Section 3.3, under the “collective classifier training”)

Step 3: FA detection by collective classification (algorithm detailed in Section 3.3., under “collective classification algorithm”

3.4. ATR Training and Other Issues

We conclude the description of FA detection, and consequently, our overall ATR system, by addressing the ATR training and other issues posed by the management.

1) For ATR training: as described previously our segmentation algorithm is unsupervised and requires no training and this is one of its advantages; our FA classification algorithm does require training but this is not unusual. What happens when a new target is added whose intensity is close to a common object (such as that of toothpaste)? For our segmentation algorithm, as long as the new target is not next to (attached to) any objects whose intensity distributions (i.e., histograms) are the same or too close to that of the new target, the new target can be detected in segmentation. For FA detection, if the new target’s other features (other than the mean intensity which is similar to some common objects), such as intensity variance, histogram features, and other features (e.g., mass and size), are not the same as those of false alarms (some of which are common objects), it will be not be labeled as FA and will therefore be detected by our ATR. On the other hand, if this is not the case, if all its features are identical to those of a common object, it might be rejected as an FA.

2) For new target types: our segmentation algorithm generally requires no changes while our FA classification algorithm may need to be retrained, if the intensity distributions of the objects in the images change significantly (e.g., when a new type of CT machine is used).

3) Use of shape in classification: shape is not used as a feature since targets may appear in any shapes.

4) How can PD and PFA be improved without over-training? Since our segmentation algorithm does not require training, over-training would not be a concern. As for FA classification, our observations suggest that most of the false alarms generated by our ATR are caused by merged objects. This happens when there are severe metal artifacts in the image and when objects with similar intensities are packed closely together. Hence, metal artifact reduction and 3D edge detection between objects with similar intensities can potentially reduce object merges and subsequently, reducing the amount of training data needed.

4. Relation to Prior Art and Innovation

We can divide the prior works relevant to our ATR system into two categories, the "academic" and the "industrial." Works in the former are usually reported in journals, such as *IEEE Trans. on IP (Image Processing)* and *IEEE Trans on PAMI (Pattern Analysis and Machine Intelligence)*. Works in the latter are usually disclosed as patents [13]. For image segmentation, there is a great deal of prior academic work and papers, and these can be divided into four main categories, namely, region growing, pixel classification/clustering, curve evolution, and graph cut [14]. In region growing, neighboring pixels whose intensity differences are small are grouped into regions. While this approach is simple to implement, it is difficult to select the right intensity difference threshold for what would be considered to be small. In curve evolution, an initial curve is evolved to stop at object boundaries. This elegant approach works well for situations where several spatially well separated objects are embedded in a background. In our application, however, the objects in the bags are often intertwine and tightly packed. Graph cut is essentially a binary segmentation approach which divides an image into a background and a foreground. In some ways, it is more restrictive than curve evolution (which can simultaneously locate several well separated objects). Based on the above considerations and our extensive observations of the project image data set (described in Section 2), we believe the pixel classification approach is more suitable for this project and consequently, this is the approach we have taken. Furthermore, our pixel classification approach is an extension of our earlier work in 2D image segmentation [7]-[10]. Finally, most prior patents used region growing for image segmentation, sometimes aided by target-specific processing, such as separating bulks from sheets. Compared to the academic prior work, these are more limited in their generality and optimality.

For data or pattern classification (for which our FA detection is a special case), there are also a great deal of academic prior work and papers. Furthermore, recent advances in machine learning has brought in some very powerful classifiers, such as the SVM, random forest, boosting, and deep neural networks. Although our use of the SVM is not novel in the academic sense, it is novel in two other aspects. First, to the best of our knowledge, our collective classification scheme is novel

Target Type	Target Subtype or Form	Level of Difficulty	Num Targets	Num Detected	PD [%]
Target	All	All	407	363	89.2
Target	Clay	All	111	105	94.6
Target	Rubber	All	158	142	89.9
Target	Saline	All	138	116	84.1
Target	Bulk	All	270	237	87.8
Target	Sheet	All	137	126	92
Target	All	Low	77	73	94.8
Target	Clay	Low	29	29	100
Target	Rubber	Low	22	22	100
Target	Saline	Low	26	22	84.6
Target	Bulk	Low	56	52	92.9
Target	Sheet	Low	21	21	100
Target	All	High	317	278	87.7
Target	Clay	High	82	76	92.7
Target	Rubber	High	125	110	88
Target	Saline	High	110	92	83.6
Target	Bulk	High	201	173	86.1
Target	Sheet	High	116	105	90.5
Pseudo-target	Sheet	High	10	8	80
Num Non-targets	Num FAs	PFA [%]			
1370	133	9.7			

Figure 9: Some Typical Results from an ATR Run: 89%PD, 10%PFA, and 80%PD for PT Sheets.

and second, most prior ATR systems as described in the patents did not use advanced classifiers, including the SVM.

5. Results and Discussions

We have tested our ATR system on the project image data set, containing 190 3D CT scanned bag images. Our target detection results were evaluated using Franco Rupicich's program, which provides PD and PFA as well as other statistics. We experimented with various parameter settings or system configurations which allow us to obtain some trade-offs between PD and PFA. Good results were obtained with segmentation parameters set to $K = 4$ (K is the number of classes), $\beta = 100$, and a 26-pixel neighborhood system. Some typical results are shown in Fig. 9, with a final 89.2% PD and 9.7% PFA, which meets the management goal of 90% PD and 10%PFA. These results also include an 80% PD for pseudo target sheets.

Case Number	Target Detected?
1	Yes
2	Yes
3	No
4	Yes
5	Yes
6	Yes
7	Yes
8	Yes
9	No
10	Yes

Figure 10: Result Summary of the 10 Cases.

Based on management request, we also singled out our results on 10 images or cases for further examination. These images are generally "difficult cases" due to, e.g., the metal artifacts in the images. What one wants to know is that whether the targets in these images are detected and if not, why not. A summary of our results is shown in Fig. 10 while, due their large volume, the detailed imagery results and commentaries are put in Appendix A. As shown in Fig. 10, the targets of interest are detected for all cases, except for cases 3 and 9. In case 3, the target was detected by segmentation but was rejected incorrectly by FA detection. In case 9, the target was not detected by segmentation since severe metal artifacts caused the target (a rubber sheet) to be split into several smaller regions and none of them meets the recall requirement.

Similarly, we also singled out two cases where our ATR system produced false alarms and these are shown in Appendix B. In both cases, a false alarm region was not detected by our FA detection algorithm because its intensity distribution is similar to that of a target. This is also due to the fact that we have set the FA classifier's pfa to a low number (around 3%) to preserve a high overall ATR system PD; as a result, the pd of the FA classifier is not very high (around 80%) and the classifier missed the false alarm regions.

Finally, we also investigated why our segmentation algorithm missed some of the targets. It turns out that in most cases, such misses were due to merging, i.e., a target is merged with a nearby object. There are two reasons for such merges: 1) the two objects have very similar intensity distributions and are next to each other and 2) metal artifacts altered the intensity distributions of the objects (making them more similar to each other than they should be). This problem will be addressed in future work (see also Section 6).

6. Conclusions and Future Work

In conclusion, our ATR system is successful and meets the management goal of 90% PD and 10% PFA. Its strengths are:

1) The system achieves good results, essentially with just two operations: pixel classification (for

segmentation) and FA detection (to obtain the final detected targets).

2) The algorithms are general and do not require metal artifact correction, bulk/sheet classification, explicit split-and-merge, shape information, and other target specific processing

3) Our segmentation and FA detection algorithms are not seen in prior patents and our collective classification technique is not seen in prior academic literature

4) Our system is relatively robust to moderate to medium metal artifacts

Our system also has some weaknesses. Specifically, in segmentation results, targets can still be merged with attaching objects of similar intensity distributions, causing them to be missed by the ATR system. Hence, for future work we plan to investigate techniques for splitting such merged targets/objects, including using 3D edge information, detecting merged regions, and re-segment regions produced by our current segmentation algorithm. Indeed, a preliminary attempt with the last approach (re-segment) produced a 94% PD (before FA detection); however, the PFA (before false detection) is still too high (in the 80% range); as a result, the final (after FA detection) PD/PFA does not improve our current PD/PFA results. Hence, more studies is need in this direction.

References

[1] C. Crawford, "Statement of work: advances in automatic target recognition (ATR) for CT-based object detection system - task order 4 and ATR initiative," 2013.

[2] D. H. Ballard and C. M. Brown, *Computer Vision*, Prentice-Hall, 1982.

[3] A. P. Dempster, N. M. Laird, and D. B. Rubin, "Maximum-likelihood from incomplete data via the EM algorithm (with discussion)," Journal of the Royal Statistical Society; Series B, Vol. 39, pp. 1-38, 1977

[4] T. K. Moon and W. C. Stirling, *Mathematical Methods and Algorithms for Signal Processing*, Prentice-Hall, 2000.

[5] R. O. Duda, P. E. Hart, and D. G. Stork, *Pattern Classification*, Wiley and Sons, 2001.

[6] R. Chellappa and A. Jain, *Markov Random Fields: Theory and Applications*, Academic Press, 1991.

[7] J. Zhang, "The mean field theory in EM procedures for Markov random fields," *IEEE Trans. ASSP*, Vol. 40, pp. 2570-2583, October, 1992.

[8] J. Zhang, "The mean field theory in EM procedures for blind Markov random field image restoration," *IEEE Trans. Image Processing*, Vol. 2, pp. 27-40, Jan., 1993.

[9] J. Zhang, J. W. Modestino, and D. Langan, "Maximum-likelihood parameter estimation for unsupervised stochastic model-based image segmentation," *IEEE Trans. Image Processing*, Vol. 3, pp. 404-420, July, 1994.

[10] J. Zhang, "The application of the Gibbs-Bogoliubov-Feynman inequality in mean field calculations for Markov random fields," *IEEE Trans. Image Processing*, Vol. 5, pp. 1208-1214, July, 1996.

[11] D. Chandler, *Introduction to Statistical Mechanics*, Oxford University Press, 1987.

[12] K. Murphy, *Machine Learning: A Probabilistic Perspective*, MIT Press 2012.

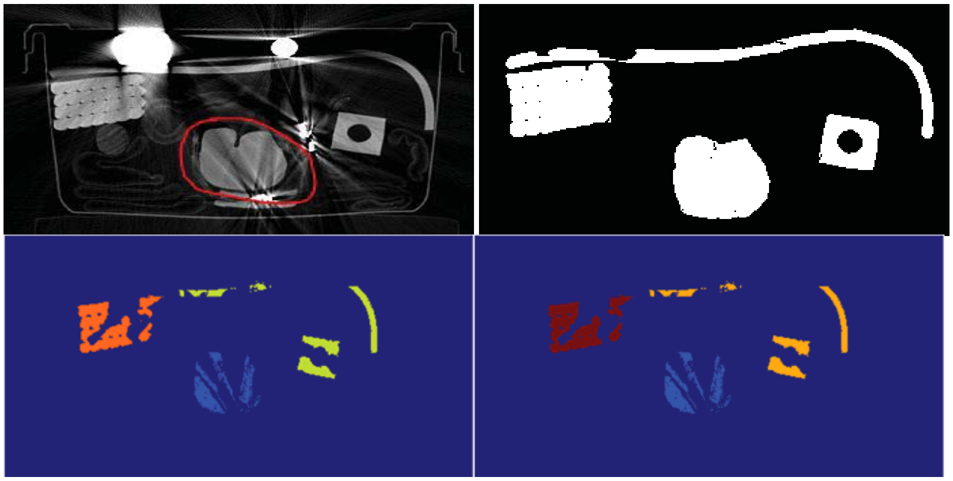
[13] C. Crawford, some ATR patents, http://www.csuptwo.com/patents.html

[14] M. Sonka, V. Hlavac, and R. Boyle, *Image Processing, Analysis, and Machine Vision*, 3rd ed, Thomson Learning, 2008, Chapters 6 and 7.

Appendix A. The 10 Cases

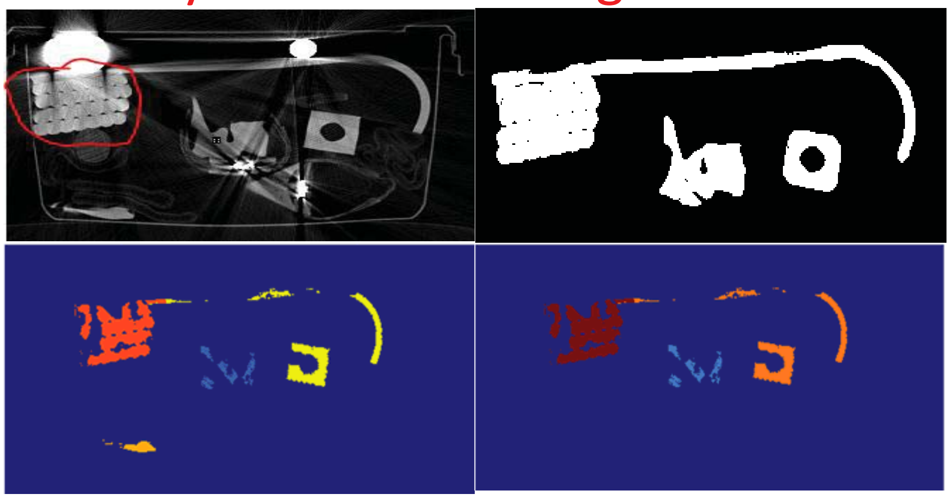
See next page

Case 1: SSN 13, Bulk with Bad Streaks



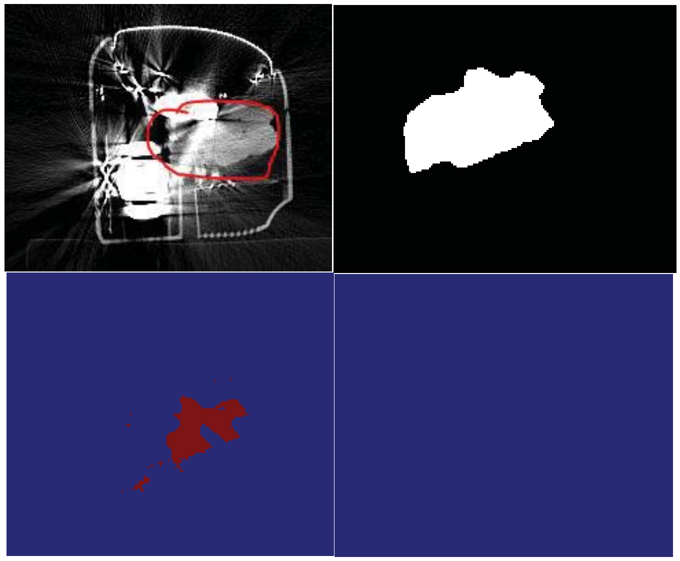
- First row: CT image and ground truth (GT)
- Second row: segmentation and after FA detection
- Target Detected? Yes.
- Detected by segmentation (with P=0.95, R=0.51; low recall due to detection window) and kept by FA detection

Case 2: SSN 13, Bulk with Bad Shading Caused by Beam Hardening and Scatter



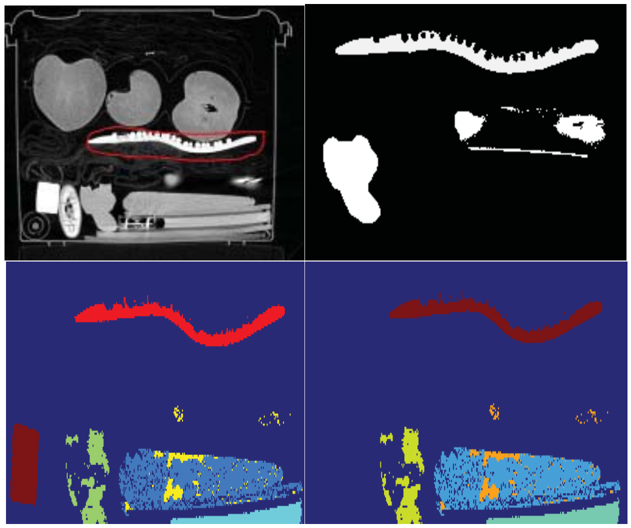
- First row: CT image and ground truth (GT)
- Second row: segmentation and after FA detection
- Target Detected? Yes.
- Detected by segmentation (with P=0.90, R=0.59; low recall due to detection window) and kept by FA detection

Case 3: SSN 35, Bulk inside Electronics



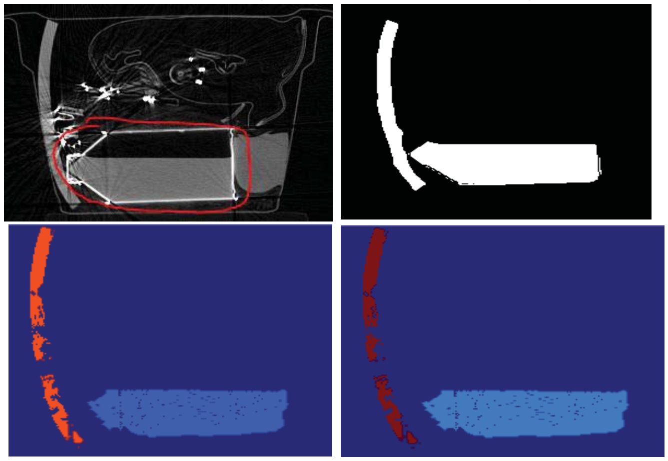
- First row: CT image and ground truth (GT)
- Second row: segmentation and after FA detection
- Target Detected? No.
- It was there after segmentation (with P=0.92, R=0.45; low recall due to metal artifacts). Our FA classifier labeled it as an FA and removed it.

Case 4: SSN 193, Bulk with Texture



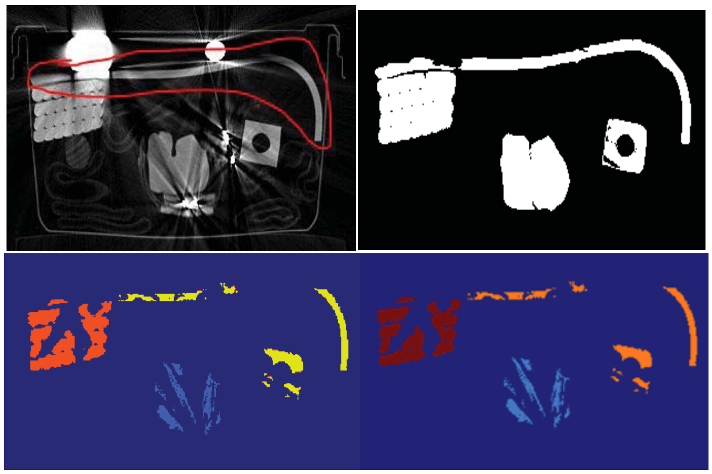
- First row: CT image and ground truth (GT)
- Second row: segmentation and after FA detection
- Target Detected? Yes.
- Detected by segmentation (with P=0.99, R=0.66) and kept by FA detection.

Case 5: SSN 63, Bulk with density close to water (~5% saline)



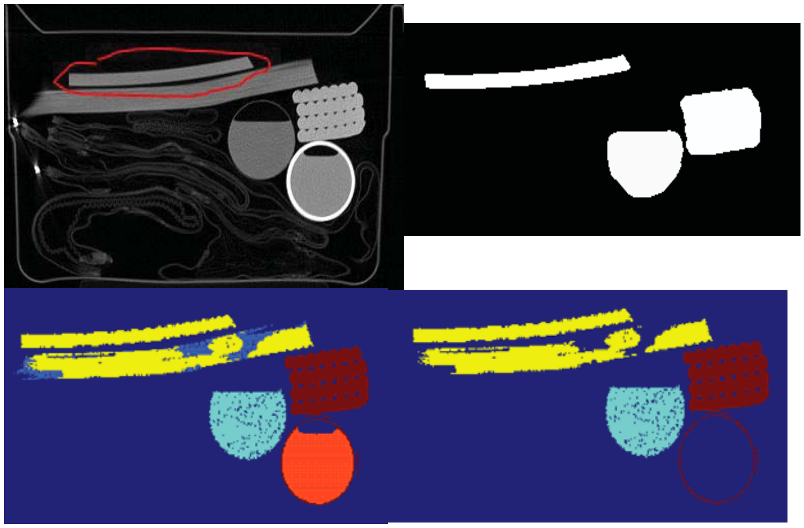
- First row: CT image and ground truth (GT)
- Second row: segmentation and after FA detection
- Target Detected? Yes.
- Detected by segmentation (with P=1.00, R=0.61) and kept by our FA detector.

Case 6: SSN 13, Sheet with bad streaks caused by metal, beam hardening and scatter



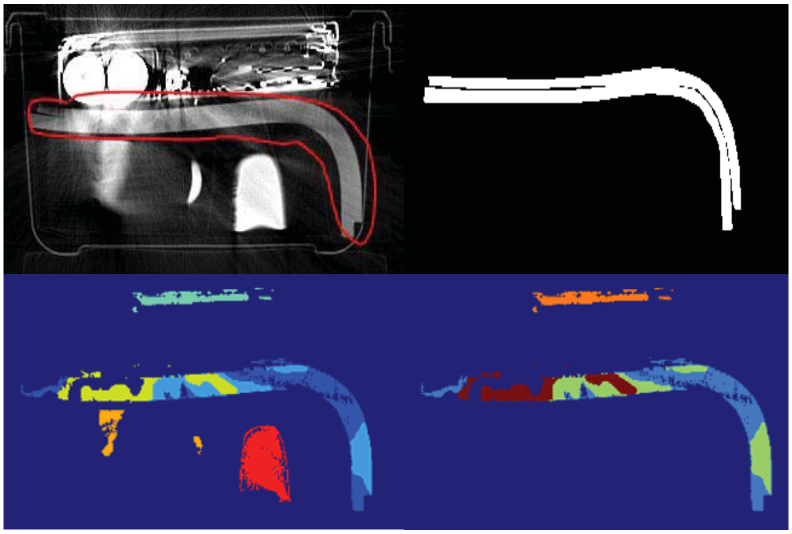
- First row: CT image and ground truth (GT)
- Second row: segmentation and after FA detection
- Target Detected? Yes.
- Detected by segmentation (with P=0.49, R=0.26) and kept by our FA detector.

Case 7: SSN 33, Sheet laying on top of another flat object



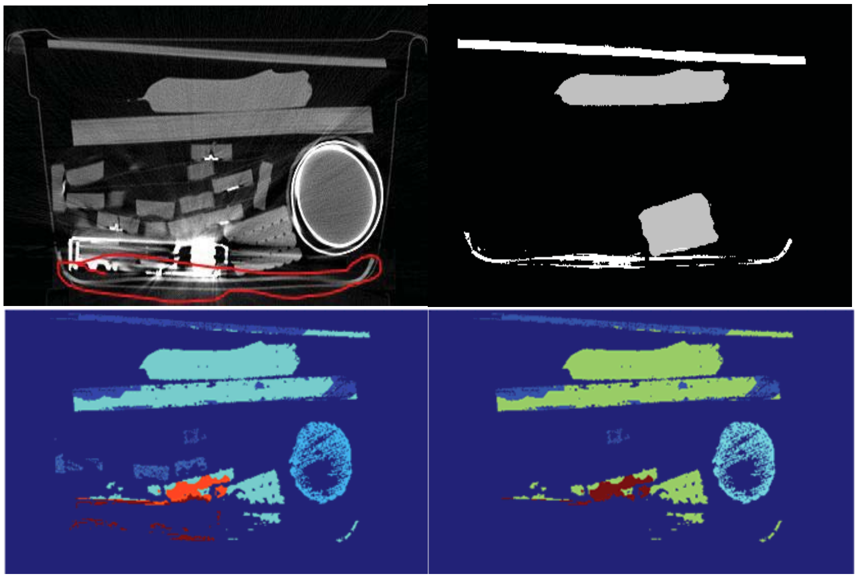
- •First row: CT image and ground truth (GT)
- •Second row: segmentation and after FA detection
- •Target Detected? Yes.
- •Detected by segmentation (with P=0.28, R=0.74) and kept by our FA detector.

Case 8: Object with lots of photon starvation



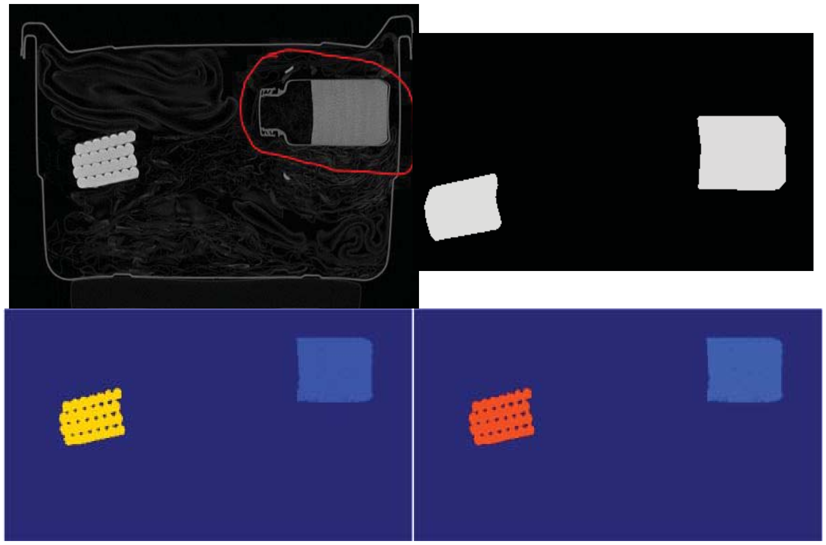
- First row: CT image and ground truth (GT)
- Second row: segmentation and after FA detection
- Target Detected? Yes.
- Detected by segmentation (with P=0.86, R=0.43) and kept by our FA detector.

Case 9: PT sheet based on thickness



- First row: CT image and ground truth (GT)
- Second row: segmentation and after FA detection
- Pseudo-targets Detected? No.
- Not detected by segmentation (with P=0.10, R=0.26), low-recall due to metal artifacts.

Case 10: PT Powder (based on density, not mass)

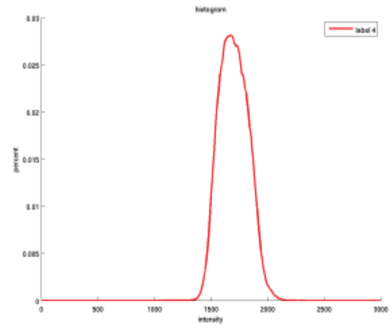


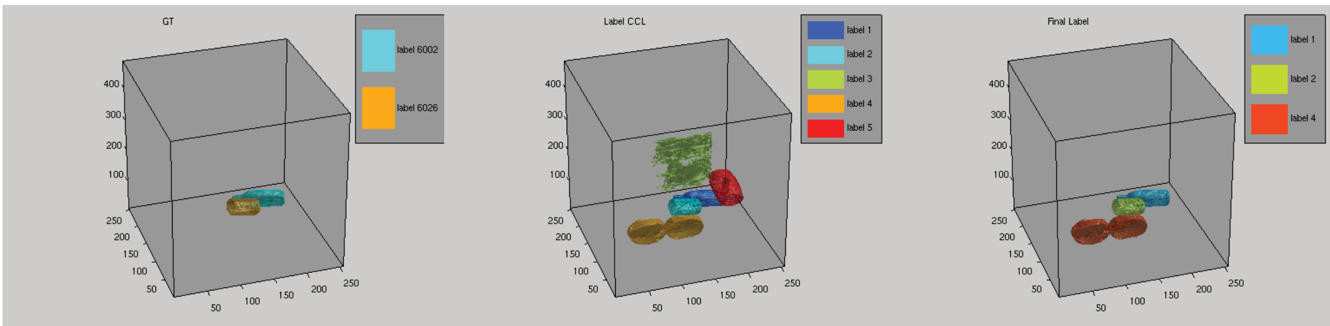
- First row: CT image and ground truth (GT)
- Second row: segmentation and after FA detection
- Pseudo-target Detected? Yes.
- Detected by segmentation (with P=0.99, R=0.93) and kept by our FA detector.

Appendix B. The 2 FA Cases

See next page

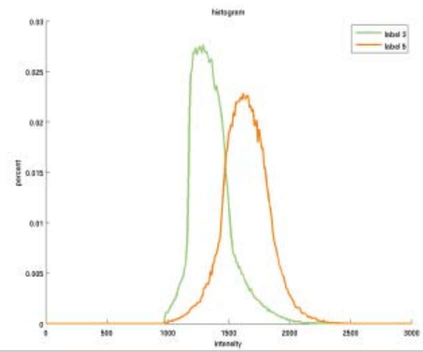
FA in SSN 008

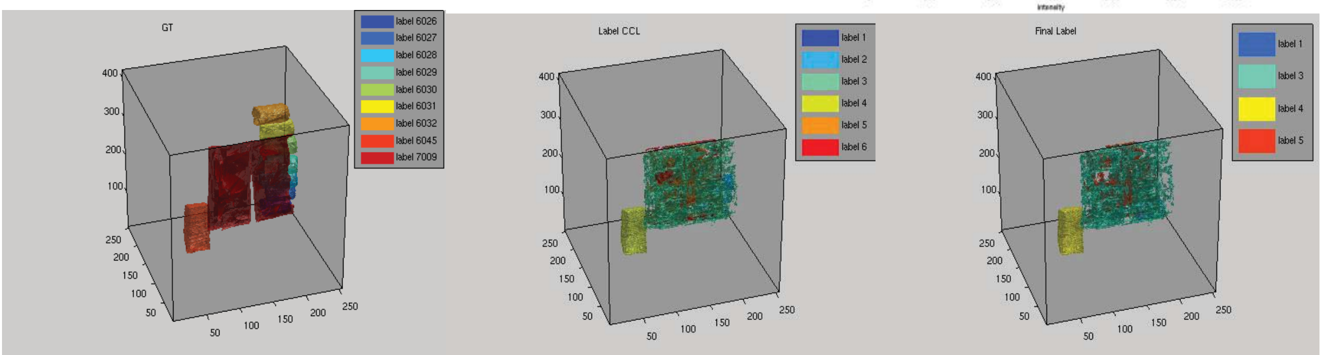




- •Top row: CT image slice for SSN 88, histogram of the FA
- •Bottom row: ground truth, ATR output, two FA objects
- •FA objects: orange
- •Observations: the FA objects have similar intensity as clay targets; our FA detection selected a pfa (to preserve high ATR PD) hence cannot detect all FAs

FAs in SSN 012





- Top row: image slice for SSN 012, histogram of the FAs
- Bottom row: ground truth, ATR output, two FA objects
- FA objects: the light green object contains some target pieces, with P=0.39, R=0.15, and not removed by our FA detector. The orange object has intensities similar to that of clay targets and was not rejected by our FA detector.

11.6.4 ATR Development – Do

"Automatic Target Recognition: Simultaneous Histogram Peak Capturing (SHPC) Technique"

Automatic Target Recognition

Simultaneous Histogram Peak Capturing (SHPC) Technique

Final Report Submitted to:
Northeastern University, ALERT DHS Center of Excellence

Submitted by:
Synho Do, PhD
Massachusetts General Hospital and Harvard Medical School

January 11, 2015

Contact:
Synho Do, PhD
Department of Radiology
Massachusetts General Hospital and Harvard Medical School
25 New Chardon Street, Suite 400B, Boston, MA, 02114
Tel: 617-643-0247
Email: sdo@nmr.mgh.harvard.edu

1

1 Contents

Table of Figures

2 Executive Summary

The project addresses a new CT-based automatic target recognition (ATR) technology development. The purpose of this document is to explain the challenges of this problem and to propose a new approach to solve the problem. ATR algorithm is known as a difficult problem because of a time limitation and guidelines of PD/PFA. And multiple vendors developed their own algorithms and a few of them are published [1] [2] but many of them are not in public domain because of security reasons. Therefore it is not easy to comprehend the current status of performances of the deployed algorithms.

Besides, the problem itself is a challenging task as we face the same problem every day in the hospital. A computed tomographic (CT) system generates 3D volume data sets and radiologists try to diagnose the disease, cancerous tissues, by looking at Hounsfield Unit (HU) values, textures, and morphological shape changes. In other words, we are trying to detect functional differences by observing structural changes in image domain. ATR is the same process we take as we do to find cancerous tissues. To detect a target, which is predetermined, we need to discover what might be the intrinsic characteristics of the target. This is an algorithm training process. The trained algorithm will be tested with real data sets for a performance evaluation and a parameter re-adjustment.

In this project, we classified three types of objects; 1) non-target, 2) target and 3) pseudo target. Also, we further classified targets to four sub-classes based on the materials (i.e., Saline, Rubber, Clay, and powder) and to two sub-classes based on different shapes (i.e., bulk and sheet). Also, containers, locations, and shapes of targets were varied arbitrarily. Therefore, the shape of target couldn't be the feature of our classifier.

The motivations of our Simultaneous Histogram Peak Capturing (SHPC) Technique are:
1) Minimum algorithm training is necessary
2) Shape independent feature selection is necessary
3) Quantitative measurement is necessary
4) Minimum volume consideration is necessary
5) Thin sheet detection algorithm to segment a defined thickness of objects is necessary.
6) Easy to add or subtract targets is necessary for a specific application
7) Easy implementation is necessary
8) Potential to be extended to multi-spectral data is necessary

In an algorithm training process, we extracted a peak HU and a HU band width (HUBW) information of the histogram for each target from the provided Ground Truth (GT) data and co-registered 3D volume data sets. The key information of the histogram of each target was stored in a look-up table for the next steps. The ATR algorithm we developed consists of four processing steps: 1) core extraction, 2) connected area decision, 3) morphological filtering, and 4) non-statistical general classifier. The details of each stage will be described in the report.

The algorithm is evaluated by using provided software. It calculates PD/PFA of labeled output of ATR (show in Table 1). In our test, our algorithm performed 100% PDs for both clay and saline detections with low difficulty and for bulk targets (All: 93.7%, Low: 98.2, High: 93) compared to sheet targets (All: 95.6%, Low: 90.5, High 96.6%, PT sheet high: 100%). For a pseudo-target sheet detection, our sheet detection algorithm segmented all targets (100% PD). Even though the ground truth image segmentation was not consistent to delineate the boundaries of thin sheet objects accurately, the algorithm could detect the target thickness of PT sheets with directional edge detector with a shift counter that can compute the thickness of target sheet. This algorithm will be explained in 4.6.4 .

The strengths of the proposed algorithm are:
1) It is easy to tune the algorithm for a specific task to find target materials. (i.e., metallic material detection, organic material detection, and liquid detection etc.)
2) No algorithm training is necessary because we use GT labels for a look-up table generation. We need to acquire an accurate look-up table for robust performance.
3) Easy to implement in the parallel computational architectures.
4) It is possible to adjust a specific target performance without changing its overall system performance.
5) It can be easily extended to multi-spectral data (i.e., data collected from dual energy or photon counting detector system) ATR (This will be discussed in section 5)

In contrast, the weaknesses of the proposed algorithm are:
1) To acquire an accurate histogram measurement or modeling is cumbersome (Repeated scans are necessary to generate an accurate parameter look-up table).

2) Ground truth image quality and mask generation process are important to keep the high quality of the look-up table (potential solutions: Single target scanning for GT labeling, Advanced image processing methods can be used for a pre-processing (i.e., Metal artifact reduction, De-noising, and Histogram sharpening etc), Iterative image reconstruction etc.)

In sum, the target preparation, scanning, and data organization were ideal for a new ATR algorithm development. The scoring tools were good enough to process multiple cases. The proposed algorithm performance could be improved by using multi-spectral imaging data in the future.

3 Introduction

The Automatic Target Recognition (ATR) [3] is a technology to identify and classify the pre-defined (i.e., Clay, Rubber, and Saline) targets from three dimensional computed tomography (CT) images. It needs to increase probability of detection (PD) and to reduce probability of false alarm (PFA) to meet the requirements of the guideline performance. However, only few of the state of art technologies developed in the companies are released in the public domain. [2, 4-7]

The purpose of this work was to develop, test, and implement an innovative ATR algorithm in an emulated environment of an airport security system. This study was developed at the Massachusetts General Hospital by means of current research trends of malignant tissue segmentation and classification [8-11]. The CT images collected from a clinical scanner (Imatron C-300) were used for the development of ATR. The in-plain pixel spacing was 0.928mm (=475/512) and the slice thickness was 1.5mm. The ground truth (GT) images and co-registered test images are provided from the ALERT DHS center of excellence.
In our test, the pseudo target detection was a challenging task with the current single energy CT data set. As shown in Table 1, our algorithm performed very well for clay and saline targets with low difficulty both 100 % PDs and for bulk targets (98.2%) compared to sheet targets (90.5%).

Table 1 PD/PFA results

Target Type	Target Subtype or Form	Level of Difficulty	Num Targets	Num Detected	PD [%]
Target	All	All	407	384	94.3
Target	Clay	All	111	104	93.7
Target	Rubber	All	158	149	94.3
Target	Saline	All	138	131	94.9
Target	Bulk	All	270	253	93.7
Target	Sheet	All	137	131	95.6
Target	All	Low	77	74	96.1
Target	Clay	Low	29	29	100
Target	Rubber	Low	22	19	86.4
Target	Saline	Low	26	26	100
Target	Bulk	Low	56	55	98.2
Target	Sheet	Low	21	19	90.5
Target	All	High	317	299	94.3
Target	Clay	High	82	75	91.5
Target	Rubber	High	125	121	96.8
Target	Saline	High	110	103	93.6
Target	Bulk	High	201	187	93
Target	Sheet	High	116	112	96.6
Pseudo-target	Sheet	High	10	10	100
Num Non-targets	Num FAs	PFA [%]			
1371	114	8.3			

4 MGH Algorithm

4.1 Requirements and Motivations

We developed a simultaneous histogram peak capturing (SHPC) technique as an ATR. The motivation of the proposed method is listed below:

- Minimum algorithm training is necessary for real applications
- Shape independent feature selection is necessary
- Quantitative measurement based algorithm development is necessary
- Minimum volume consideration is necessary
- Simple method to add or subtract targets materials is necessary
- Easy implementation is optional
- Potential to be extended to multi-spectral data ATR is necessary

Besides, it is indispensable to investigate priori knowledge of data sets and the system that we will use for the task. If we integrate our priori knowledge in the algorithm development, the algorithm itself grows into a regularized method without explicit formulations. Eventually, this is a practical approach to maximize the efficiency of the algorithm in the case we use an emulated data set.

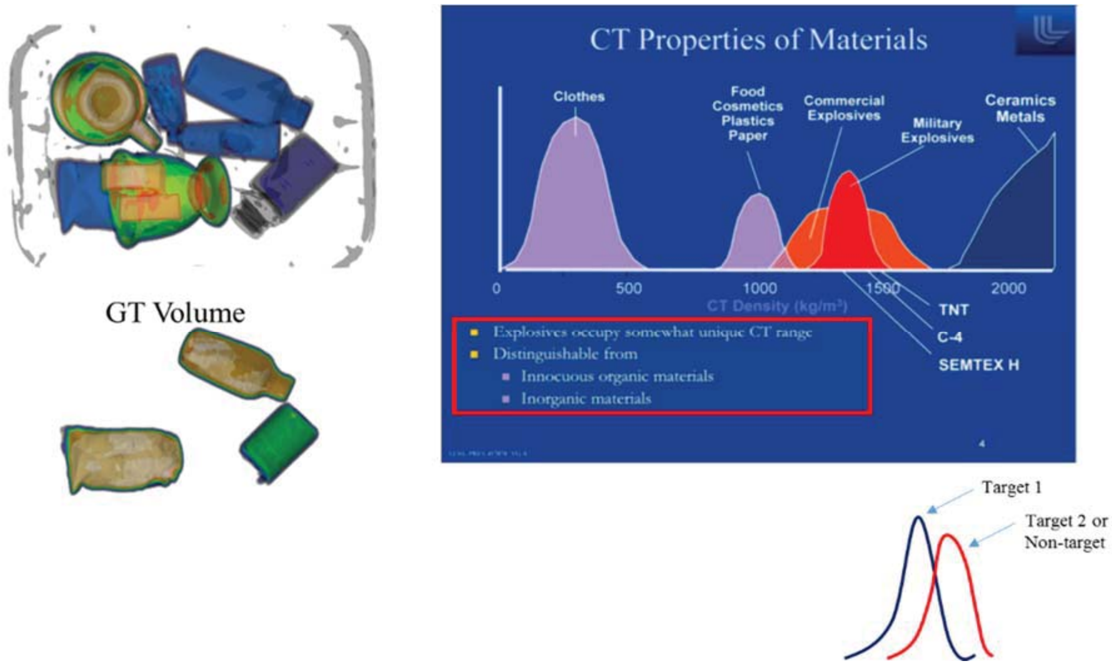


Figure 1 CT properties

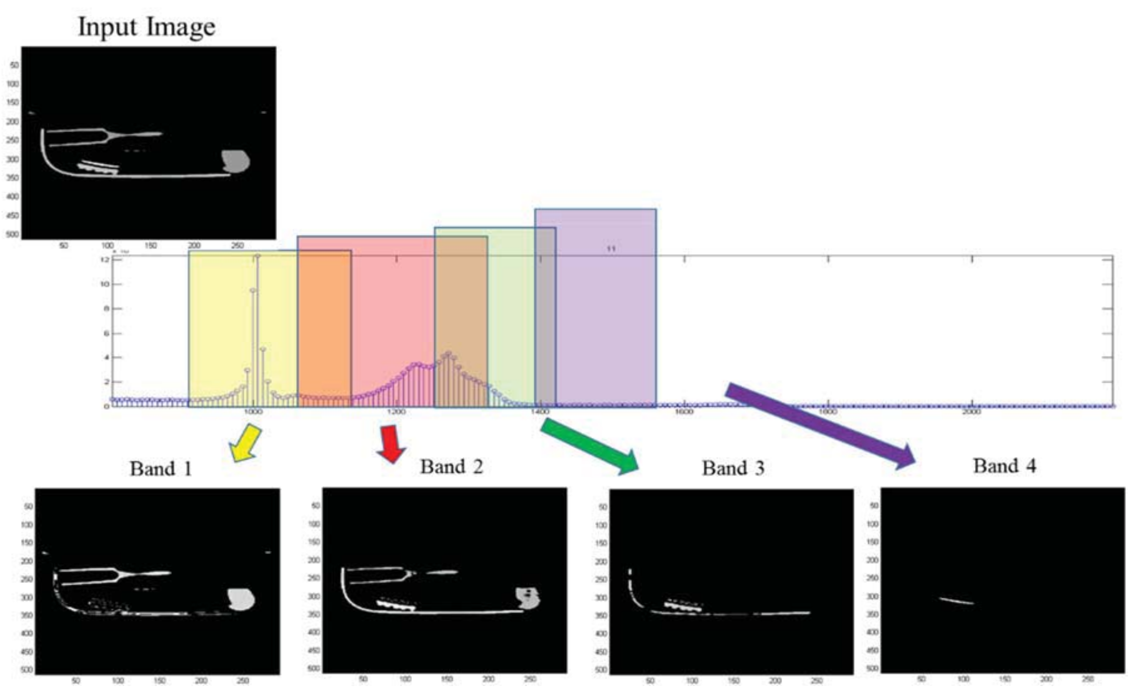


Figure 2 Conventional methods (i.e., Filter Banks)

7

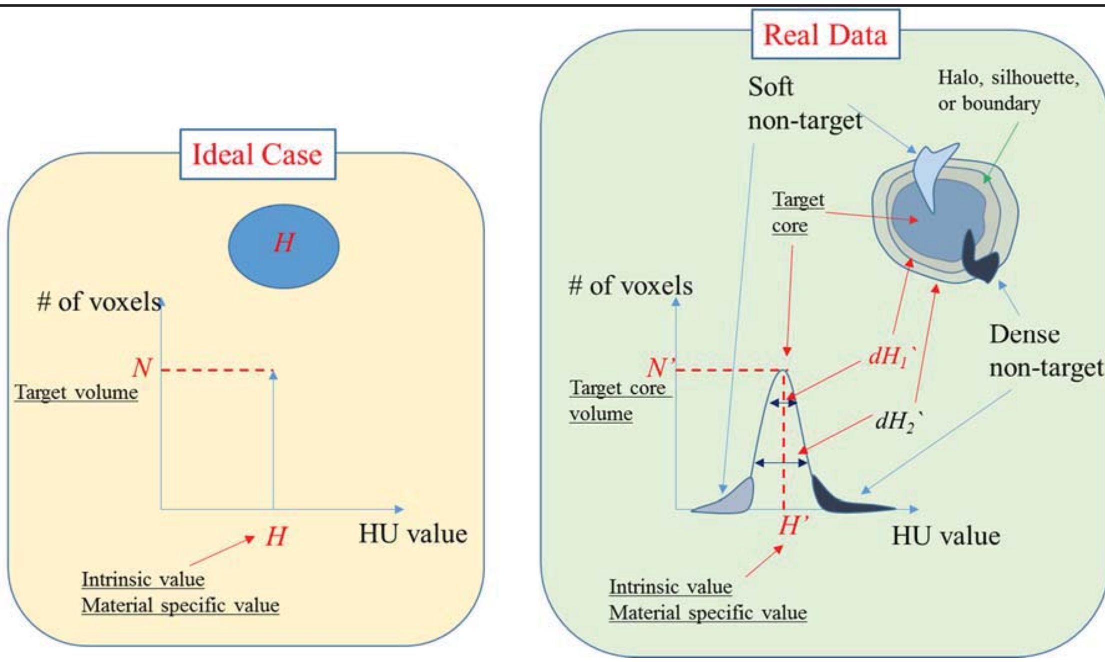


Figure 3 Ideal vs. real histogram

Our knowledge discovery process is summarized:

- CT properties (shown in Figure 1)
  - What we know about our targets:
    - Quantitative measurement (i.e., Unique HU values, voxel size, and volume etc.)
    - Target histogram characteristics (Peak location, width, slope etc.)
    - Artifacts are present (i.e., Noise, PSF, metal artifacts, and neighboring objects etc.)
  - Questions
    - What if two target histograms are placed closely to each other? ➔ Potential solution: Pre-processing, Histogram sharpening
    - It is a challenging task to estimate an accurate histogram model for each target. And each histogram shape is different and it is system dependent ➔ In other words, it is very hard to make probabilistic (statistical) predictions.
- Conventional methods (shown in Figure 2)
  - Filter Banks are efficient for a target signal detection and amplification (i.e., multi-dimensional analysis, high-order statistics analysis etc.)
  - Transform domain filter banks are simple and fast (i.e., FFT, DCT, Wavelet etc.)
  - Question
    - Can we develop an adaptive and accurate target signal detection method? ➔ If we could measure a few key parameters of a target histogram accurately from the ground truth (i.e., training data set), we can extract target signal precisely (the core of an object as shown in Figure 2)
- Ideal vs. real histogram comparison (shown in Figure 3)

- *H≈H`*: Intrinsic HU values from ideal and real data are very close to each other (CT property).
- *N≥N`*: Target core size/volume in the real data is smaller than the size/volume of ideal target.
- *dH`* : HU bandwidth (HUBW) can be determined by the characteristics of each target: (i.e., Boundary, Neighboring objects, Core volume, Merge/ Separation of objects etc.)
- Soft/Dense non-target volume : Tails of Histogram ➔ Or texture

We considered the requirements of ATR and questioned our motivations of algorithm design in order to develop an innovative technology.

4.2 Algorithm Overview

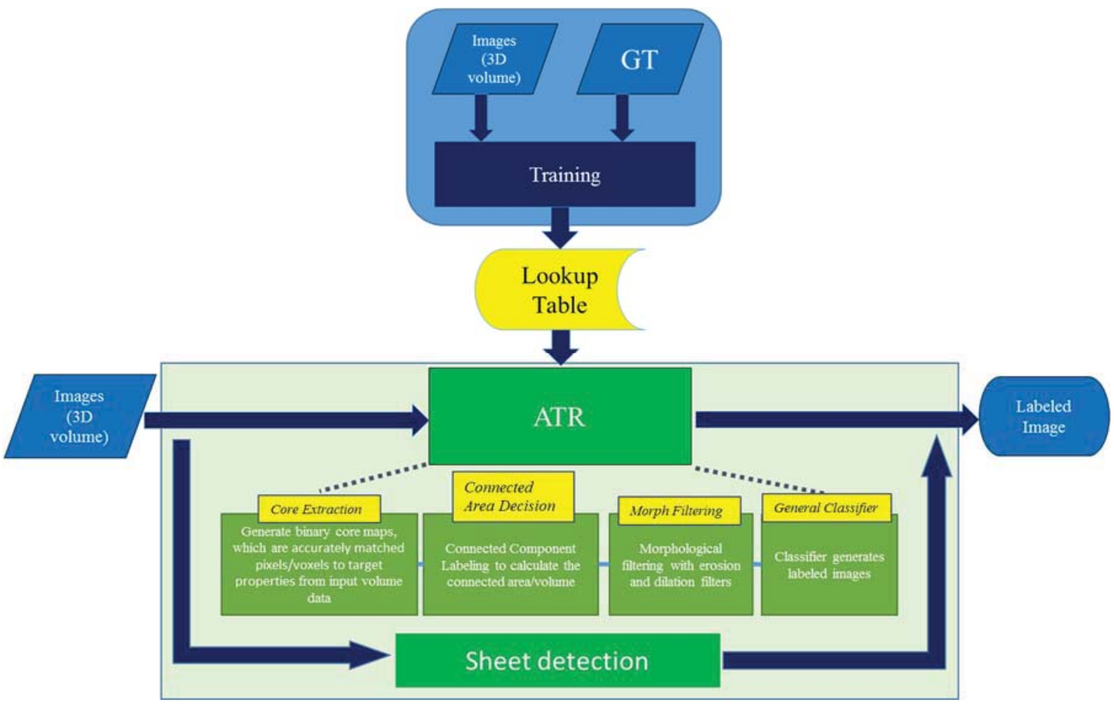


Figure 4 Algorithm Flow Chart

A training process is to extract key information of histograms for targets from the provided Ground Truth (GT) data and the co-registered 3D volume data sets as shown in Figure 4. This process for each target is single path process and it updates the lookup table, which we will use for the test. Sheet detection is performed separately in parallel.

- <u>*Core extraction*</u>: It is the first step of the proposed method to find the core (accurately matched pixels of voxels to target properties) of target. It uses a material specific HU value, which is measured from a nominal target histogram, and a HU bandwidth (HUBW), which can be determined by users.

- *Connected area detection*: To make the core extraction more robust, we remove non-core voxels by using a connected area detection process. A pre-defined threshold value is used to keep the volumes only satisfying the minimum volume requirement.
- *Morphological filtering*: The morphological erosion filter cleans up residuals of previous processing steps and the dilation filter can restore the size of target volume.
- *Target histogram*: A target histogram for a specific target can be measured by using multiple cases in the ground truth data set (like our case) or from repeated target scans. The measured target histogram will be saved as a look-up table so that we could use it for test cases.
- *Peak detection*: To extract the core of a target, we measure the HU value that gives the maximum bin count in the histogram.
- *Bandwidth selection*: The selection of HUBW is related to how much deviation of HU we could allow to find the core of a target reliably.

4.3 SHPC Simulations: (HUBW parameter selections)

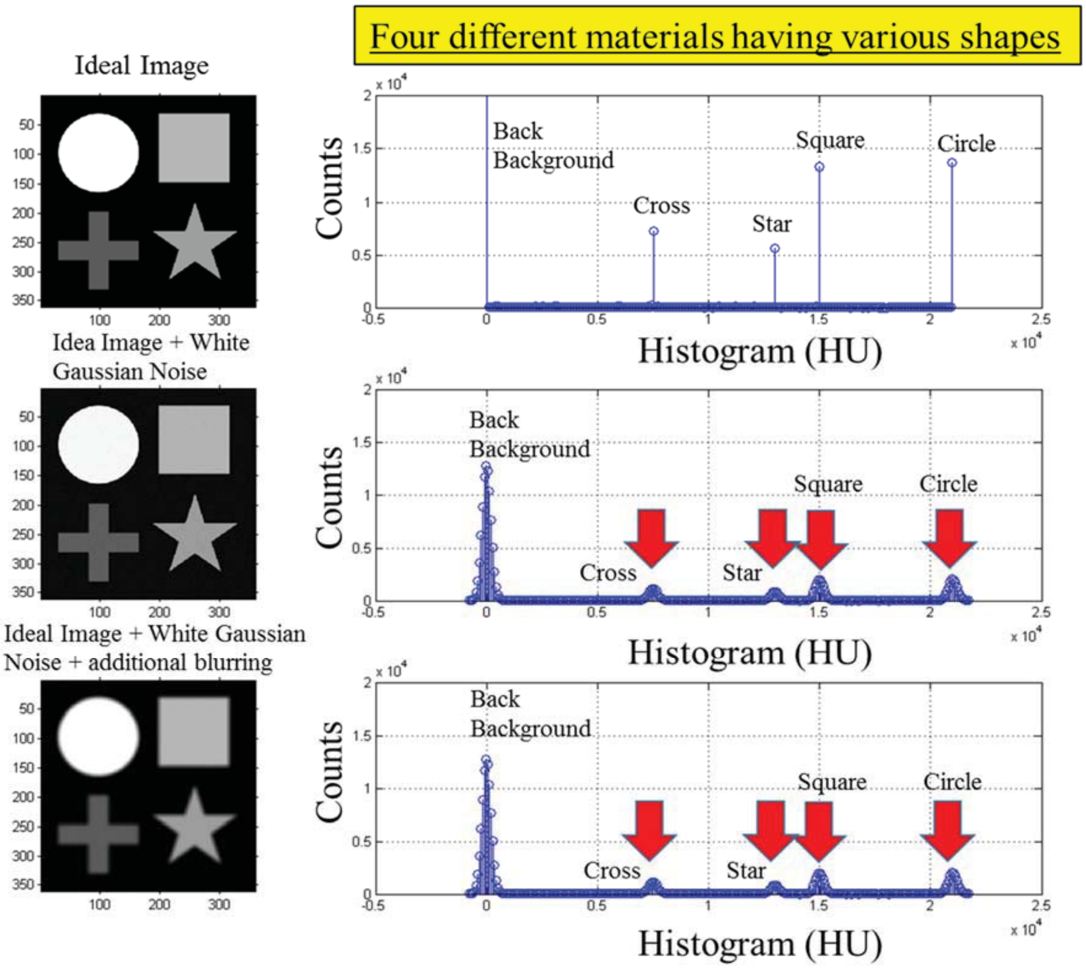


Figure 5 SHPC simulation

10

In this simulation, we generated four different shapes made with different materials.

- Each bin location is related to material characteristics of each target.
- Each bin height is the area/volume of each target.
- In noisy and blurred cases, the targets provides the same intrinsic peak locations but the heights of peaks are different from those of the ideal case.

By changing HUBW from 60 to 100, we could observer how our algorithm works in Figure 6 and Figure 7. Also, we show how objects merging and splitting can be controlled by the morphological filters in Figure 8.

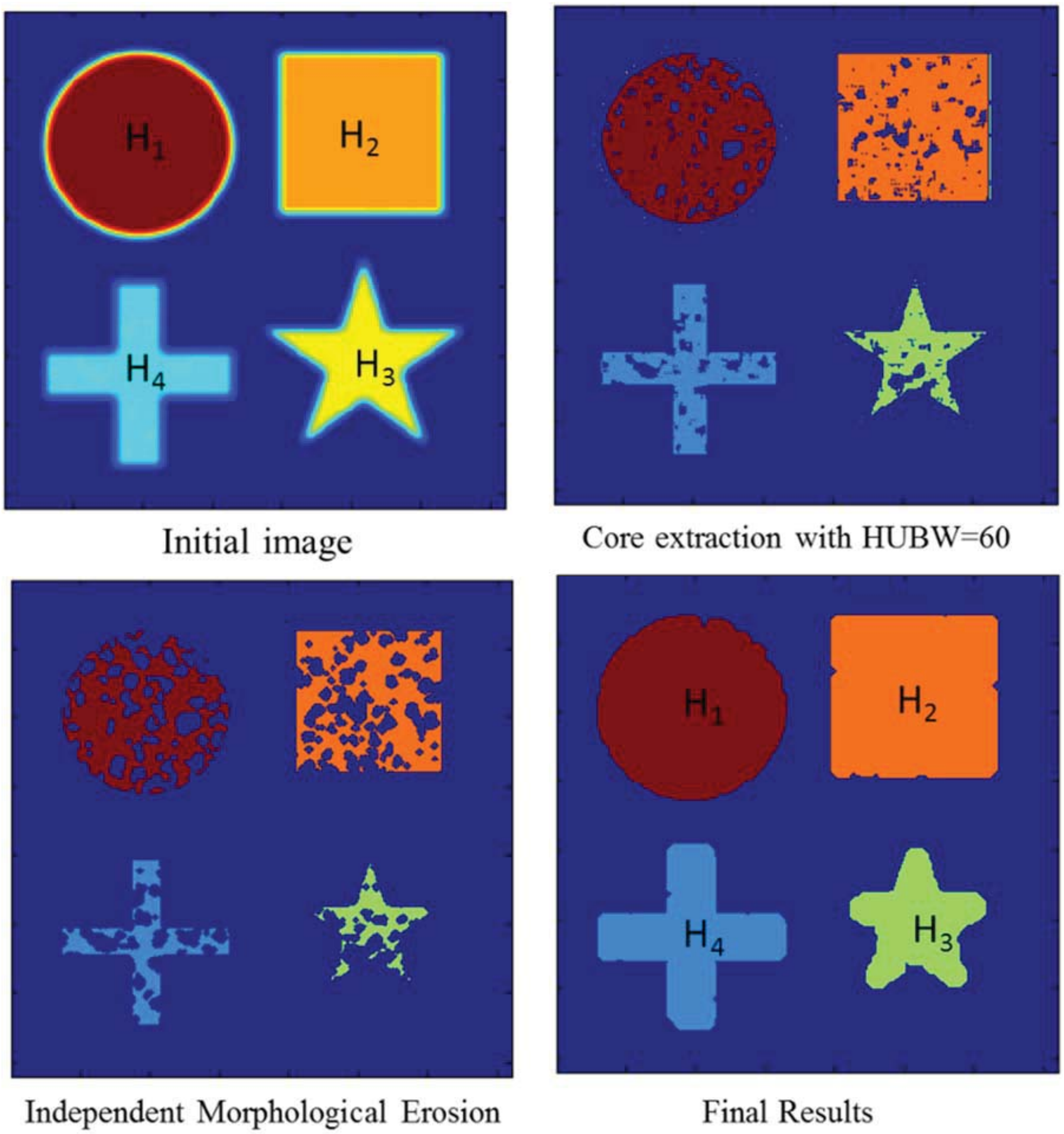


Figure 6 SHPC Simulation with HUBW=60

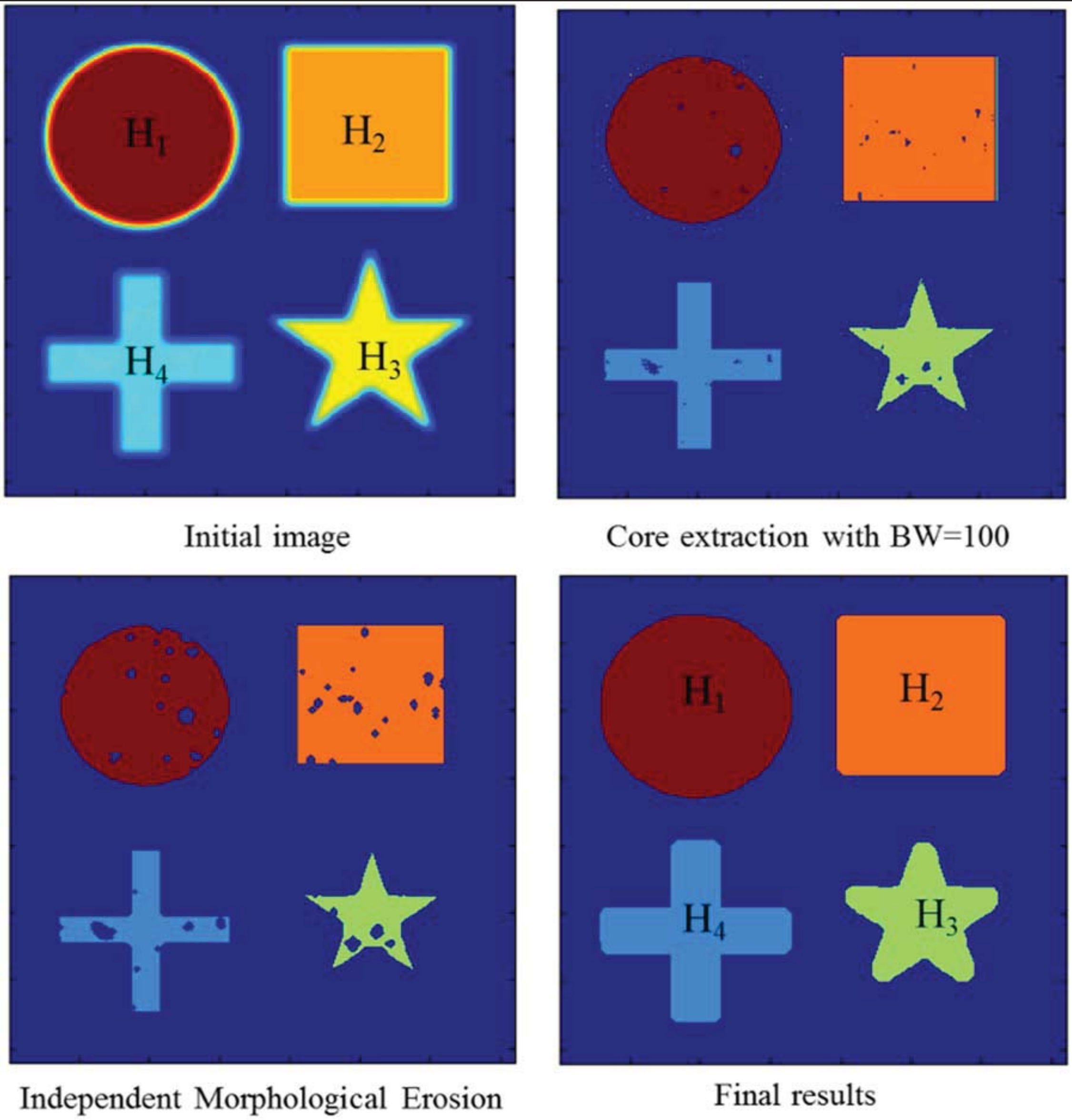


Figure 7 SHPC Simulation with HUBW=100

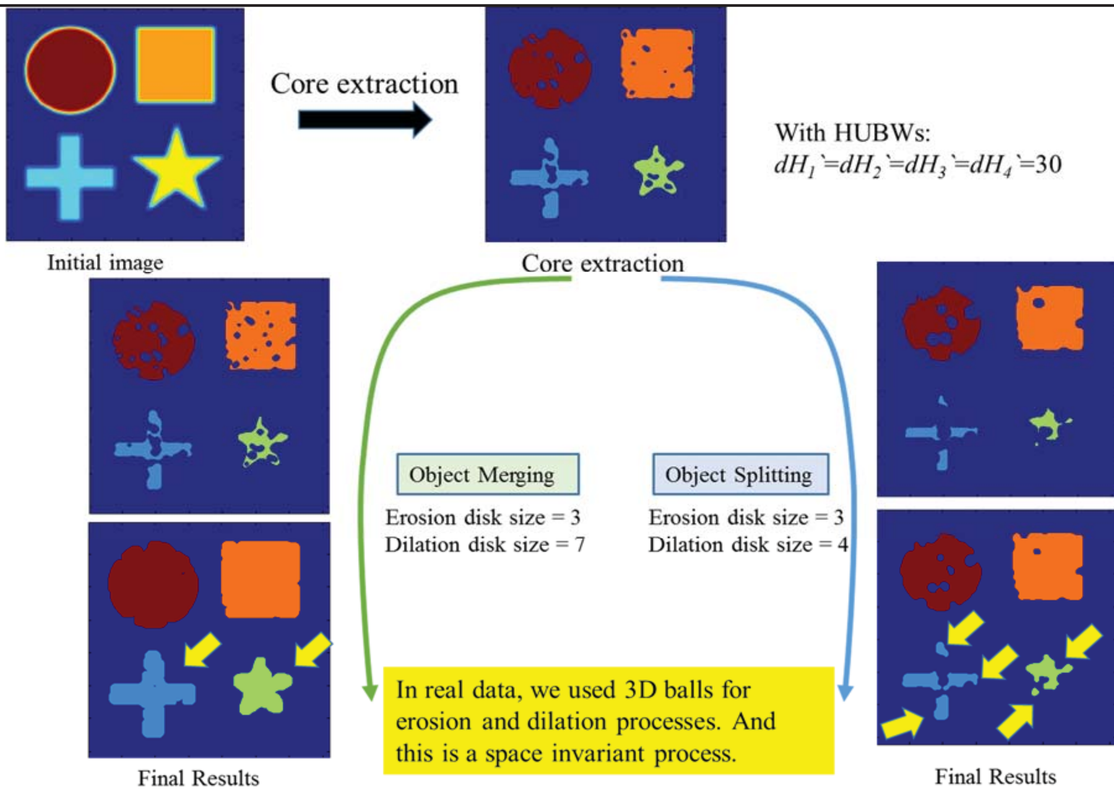


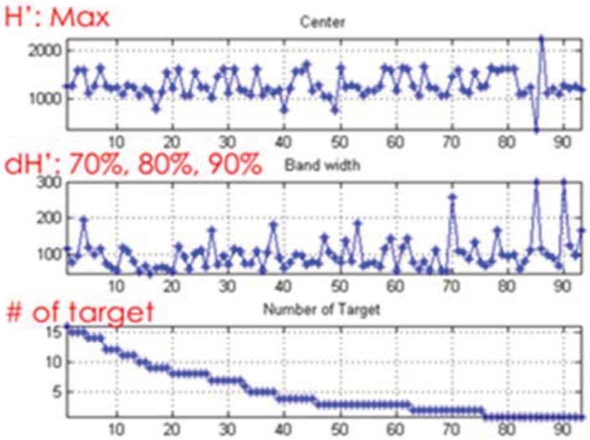
Figure 8 SHPC Simulation, Object merging and splitting

4.4 Look-up Table Parameter Generation

We processed the provided ground truth data set for a look-up table parameter generation.

- Extracted a list of target ID numbers for each scan and re-ordered scan numbers by each specific target (Figure 9)
- Extract identical targets from all scans by using the ground truth masks (3D volume data sets, Figure 10)
- Histogram sharpening: 3D median filter was implemented to improve the image quality that results in sharp histogram. (Figure 11)
- For each target, we measured a mean histogram from multiple identical targets to detect an accurate peak and to calculate HUBWs (Figure 12).

Target frequency analysis



Target	count		Target	count		Target	count
6018	16		6019	4		6192	1
2093	15		6032	4		6193	1
6156	15		6044	4		6569	1
6169	15		6088	4		6575	1
6012	14		6570	4		6595	1
6141	14		6576	4		6606	1
6150	14		6586	4	46	7006	1
6140	12		2090	3		7008	1
6157	12		6027	3		7009	1
6171	12		6029	3		7011	1
6001	11		6030	3		7012	1
6143	11		6050	3		7013	1
6172	11		6105	3		7014	1
6002	10		6142	3		8026	1
6144	10	15	6145	3		8029	1
6011	9		6177	3		8030	1
6031	9		6562	3		8031	1
6042	9		6564	3			
6046	9		6577	3			
6025	8		6592	3			
6045	8		6593	3			
6178	8		6594	3			
6179	8		6605	3			
6507	8		6615	3	63		
6517	8		2092	2			
6568	8	26	6040	2			
6028	7		6041	2			
6047	7		6146	2			
6081	7		6162	2			
6163	7		6181	2			
6506	7		6182	2			
6560	7		6512	2			
6557	6		6518	2			
6026	5		6582	2			
6051	5		6587	2			
6175	5		6590	2			
6573	5		7007	2			
38 → 8024	5		6053	1			

Figure 9 Target ID number extraction and ordering

TARGET: 6018
(Case #:10,**13**,29,34,**59**,**67**,**73**,80,86,94,102,**110**,118,**140**,157,165)

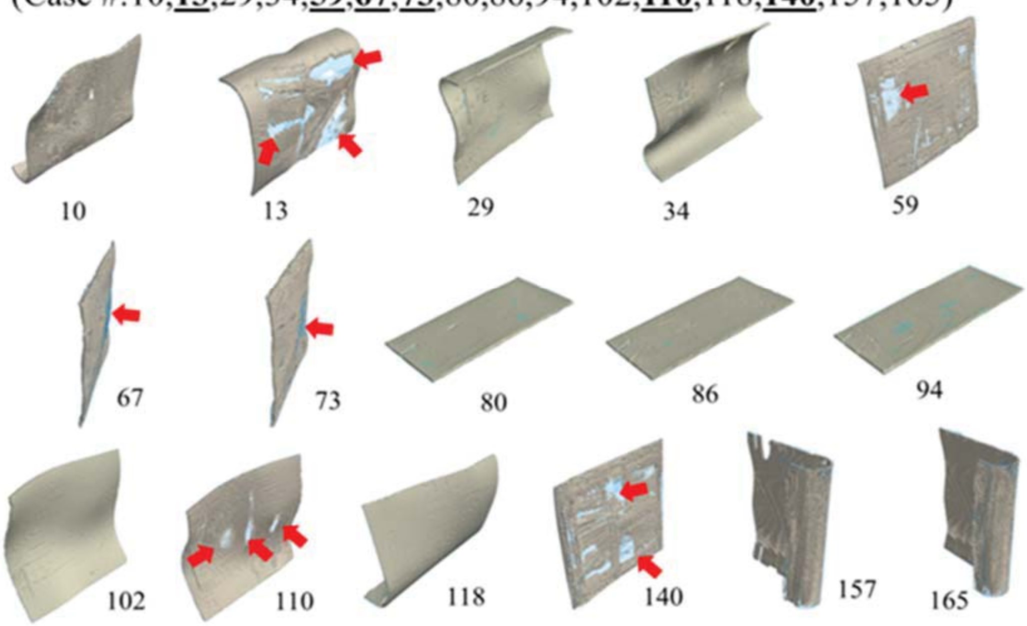


Figure 10 Identical target extraction

14

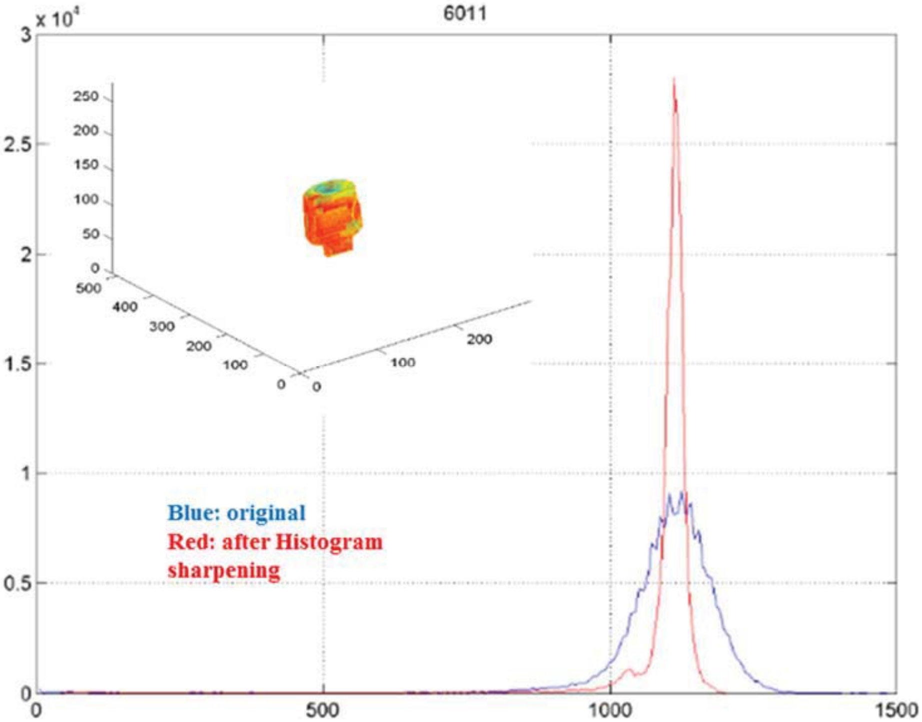


Figure 11 Histogram Sharpening

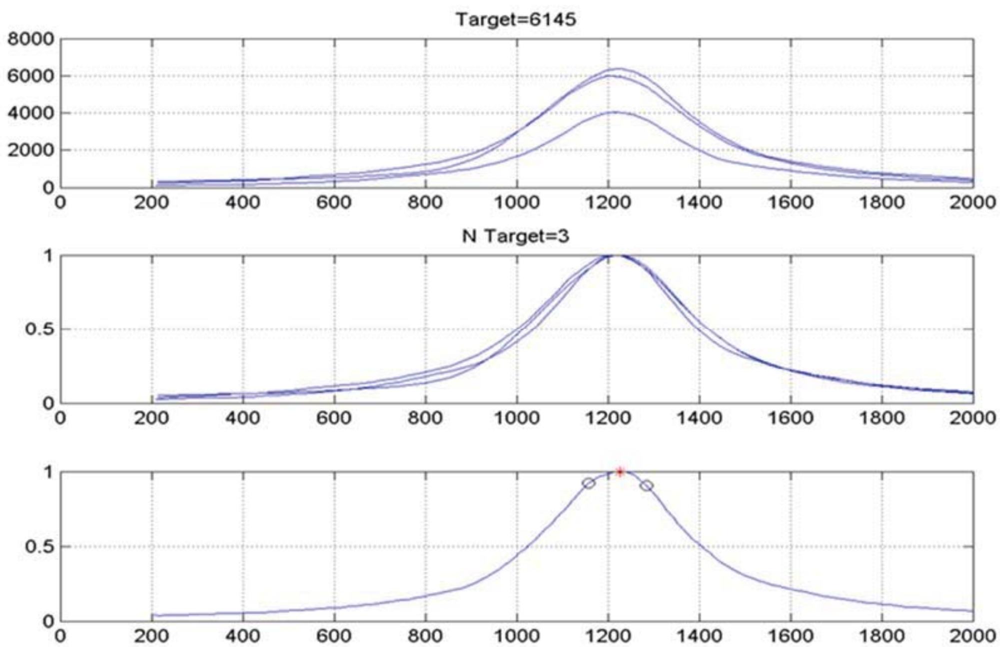


Figure 12 Intrinsic target histogram measurement

15

4.5 ROC Analysis

We tested the performance of the proposed algorithm for 188 cases of 3D volume data by changing the number of targets keeping HUBW. In Figure 13, we can observer rubber targets need only a few bins. But more bins are required for clay targets to reach >90% PD. When we use wider HUBW, it becomes easier to detect 'sheet' objects. This makes sense because sheet targets have smaller core areas compared to bulk targets as shown in Figure 13.

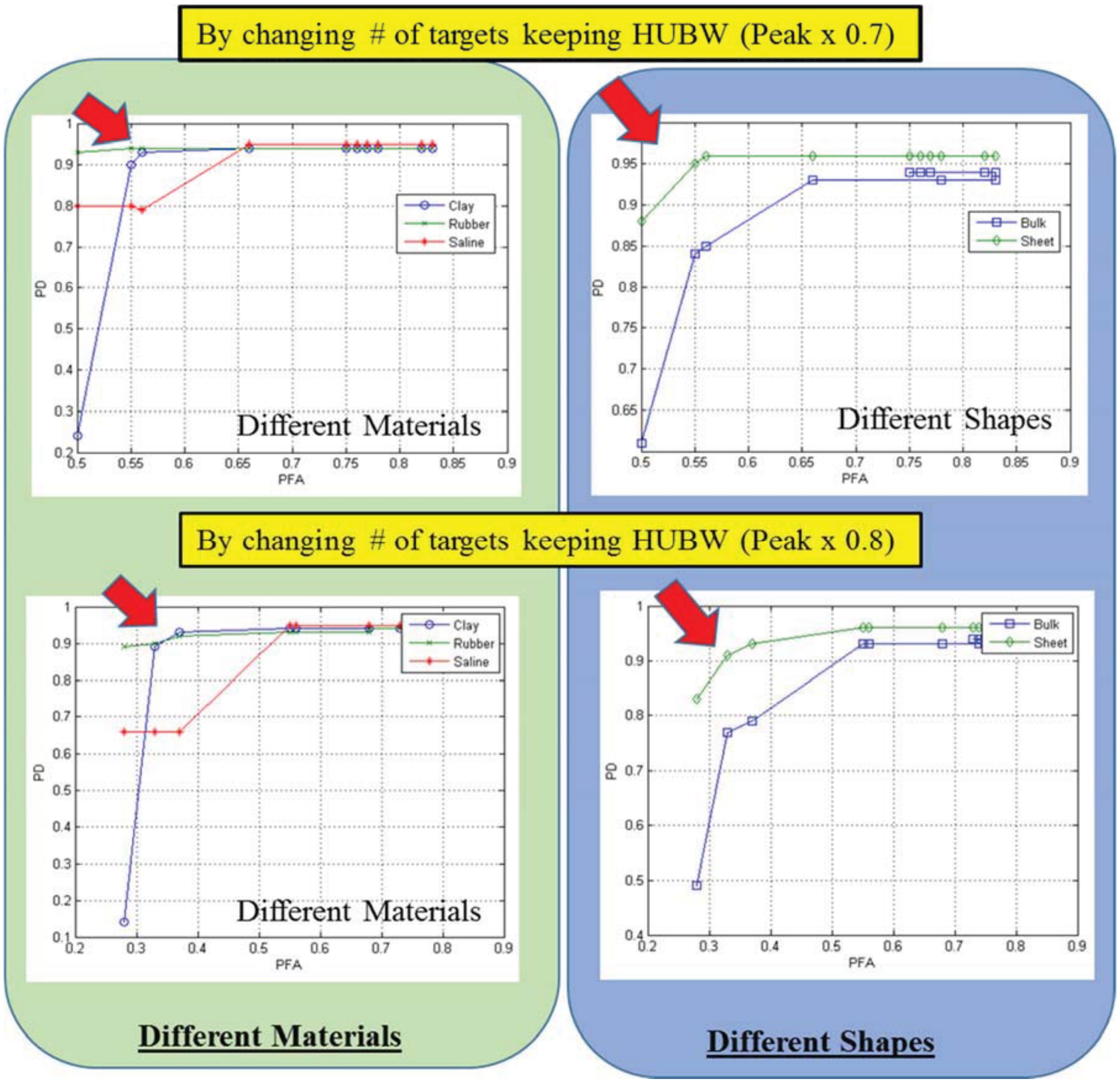


Figure 13 ROC analysis

4.6 Results

4.6.1 Single target in 3D volume

When we observe 2D image only at slice = 94, it looks as if the target was not detected. A photon starvation artifacts caused HU value change near slice = 94 so there is missing part of rubber sheet.

However we could observe that the target at slice = 180 appears detected accurately. Actually, the merged rubber target is detected correctly when we observe the results in 3D volume.

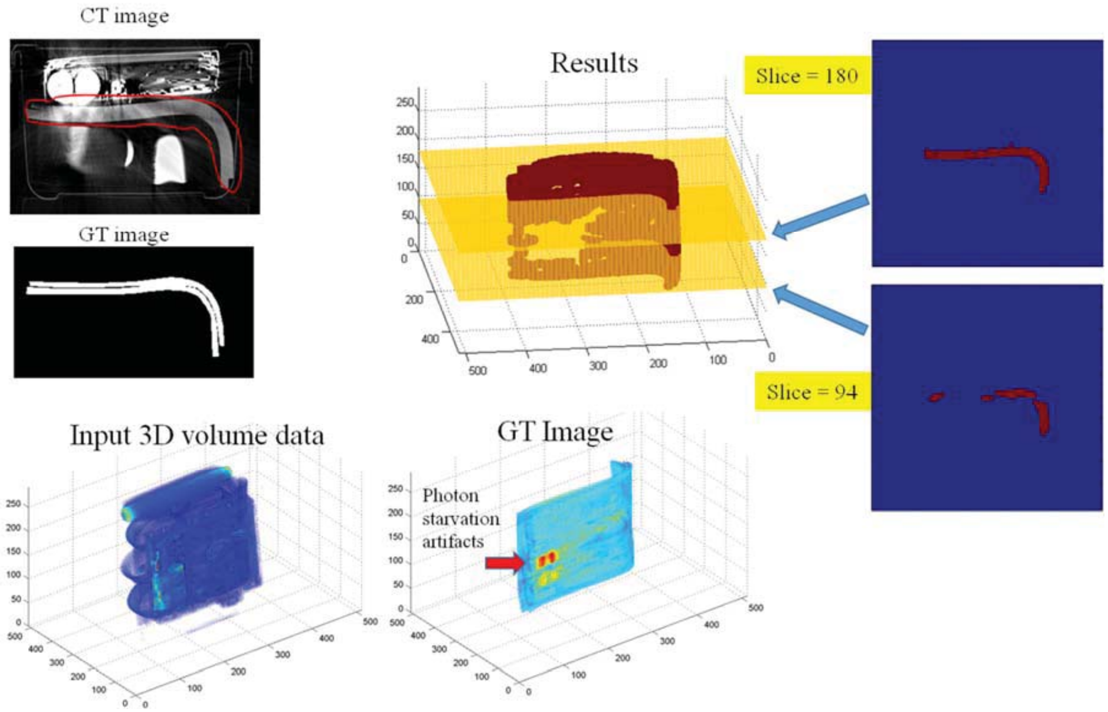


Figure 14 Presentation case 8 (Merged rubber)

4.6.2 Multiple targets in 3D volume

Clay was merged with a rubber sheet but 10 % saline is segmented accurately even in the situation with streak artifacts. The 3D volume visualization is helpful to observe segmented targets all together.

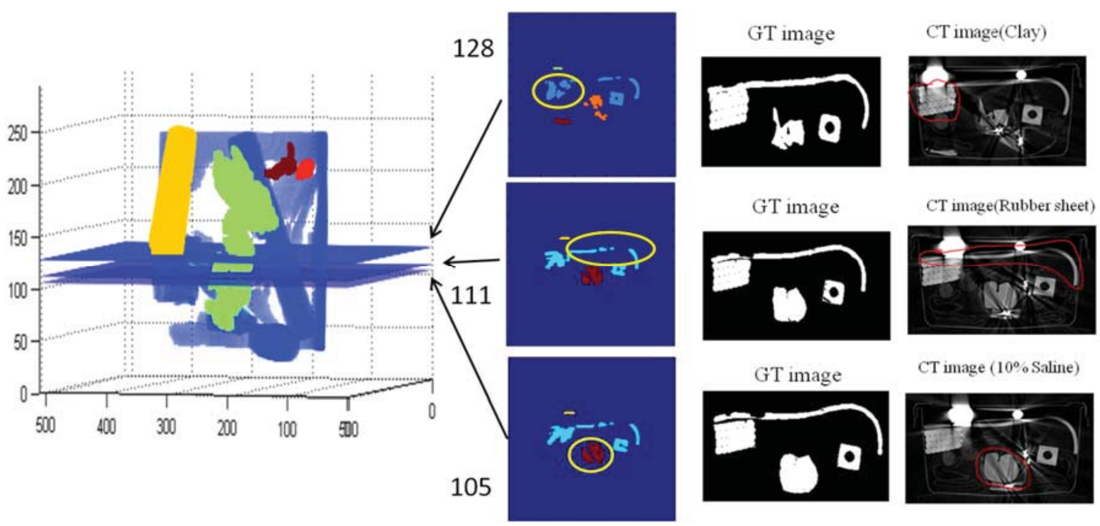


Figure 15 Presentation case (1, 2, and 6)

4.6.3 Target specific tuning

This presentation case shows how target specific tuning can be executed for a rubber sheet without changing any parameters for the saline target. The HUBW was increased (from left to right: 70, 90, 110, and 130 as in Figure 16). With 70, 90, and 110 HUBW values, the rubber sheet was not segmented as one target. However, the rubber sheet was segmented as single target with HUBW=130 without changing any performance for saline target detection.

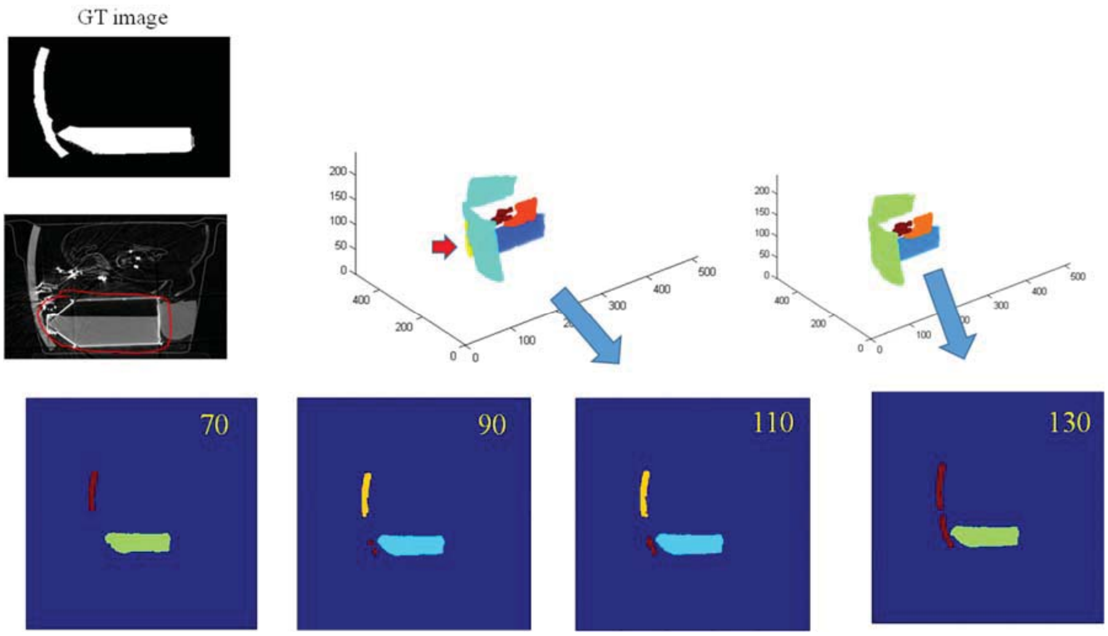


Figure 16 Presentation case (5% saline in tin bottle)

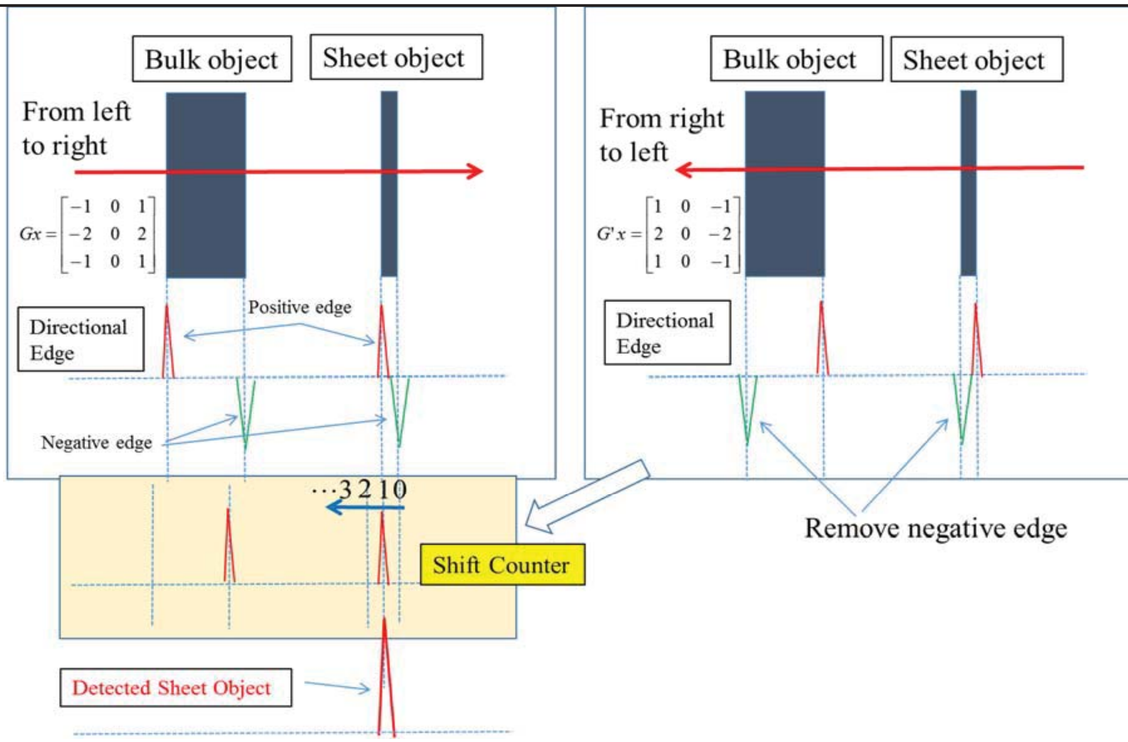


Figure 17. Directional Edge Detector with a Shift Counter

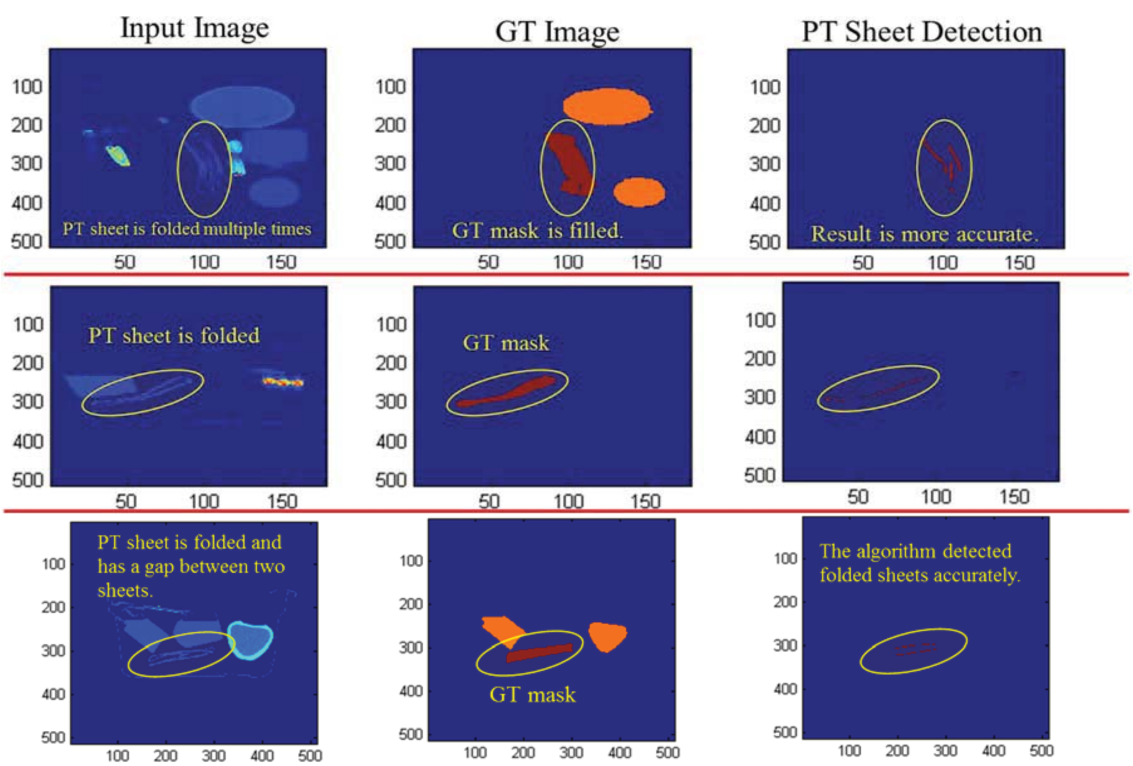


Figure 18. PT sheet detection results

4.6.4 Sheet detection: Directional edge detector with a shift counter

PT sheets are thin so those occupy only a few pixels in the 3D image domain. As shown in Figure 17, we use a directional edge detector with a shift counter to calculate the thickness of sheet objects. This algorithm can detect very thin sheet object efficiently because we have a few control parameters in the algorithm. In this algorithm, we consider:

- Material: we apply material specific HUBW filtering after histogram sharpening process
- Location: the directional edge filtering algorithm is repeated for all three planes (i.e., axial, coronal, and sagittal planes) to find a sheet structure in 3D space
- Thickness: The algorithm computes and saves the number of pixels with a shift counter when we compare the positive edges from two different directional edge maps. This process is repeated for vertical and horizontal directions for a plane therefore we need four directional edge maps for a plane.

For a plane, we calculate four directional edge maps with four 3 x 3 convolution kernel Sobel edge detectors in a matrix form because it is easy to change the filtering directions by rotating the matrix:

- From left to right edge: $\text{Gx} = \begin{bmatrix} -1 & 0 & 1 \\ -2 & 0 & 2 \\ -1 & 0 & 1 \end{bmatrix}$
- From right to left edge: $\text{Gx}' = \begin{bmatrix} 1 & 0 & -1 \\ 2 & 0 & -2 \\ 1 & 0 & -1 \end{bmatrix}$
- From up to bottom edge: $\text{Gy} = \begin{bmatrix} -1 & -2 & -1 \\ 0 & 0 & 0 \\ 1 & 2 & 1 \end{bmatrix}$
- From bottom to up edge: $\text{Gy}' = \begin{bmatrix} 1 & 2 & 1 \\ 0 & 0 & 0 \\ -1 & -2 & -1 \end{bmatrix}$

After we apply G_x and $G_{x'}$ on image f(x), it becomes:

$$f_{Gx}(x) = G_x * f(x) \tag{1}$$

Then we compute matched positive edges by using G_x and $G_{x'}$ maps (i.e., f_{Gx} and $f_{Gx'}$) and a shift counter, which compute matched edge locations with:

$$E_{Gx}(n) = \begin{cases} 1, & if \;\; f_{Gx}(x) = f_{Gx'}(x-n) \\ 0, & otherwise \end{cases} \tag{2}$$

where $1 \leq \text{n} \leq \text{N}_{set}$. N_{set} can be adjusted to the target thickness of sheet objects. For example, we used $N_{set} = 4$ in our test case. If we increase the target thickness, we can detect thicker sheet targets (i.e., $N_{set} = 8$). In addition, if we are looking for a specific sheet object with a known thickness, we can set $\text{n} = N_{set}$ to detect the sheet having $\lfloor N_{set}/2 \rfloor$ thickness. In this case, we should know the specific target thickness (i.e., 2 mm sheet). We usually put the even number for N_{set} but it is also acceptable to use odd numbers with a rounding operation (i.e., Rounding operator: $\lfloor \quad \rfloor$). In all cases, we consider the pixel spacing to calculate the absolute thickness of target object. (i.e., 2 pixels = 2 x 0.928mm = 1.856 mm)

With $N_{set} = 4$, we shift one of edge map (i.e., $f_{Gx'}$) to compare with the other edge map (i.e., f_{Gx}). By doing it, we can discover the aligned two positive edges as shown in Figure 17. In this case, there are 4 comparison processes ($N_{set} = 4$) by shifting the edge map ($f_{Gx'}$) toward the direction of edge detection process (from right to left). All matched edge pixels will be labeled as candidates for thin objects. It is an "union process" to collect all the satisfied pixels such as in Eq. (1).

$$E_{Gx}(1) \cup E_{Gx}(2) \cup E_{Gx}(3) \cup E_{Gx}(4) \tag{3}$$

This process is repeated for f_{Gy} and $f_{Gy'}$ maps to detect vertical edges since horizontal and vertical processing blocks are identical to each other except edge detection directions. This can be executed by the same function with rotated inputs.

For each plane, the horizontal and vertical processes were mainly applied. However, the diagonal edges were also detected at the same time because we used the range of *n*. When we want to detect 1-pixel thickness sheet structure, the length of edge distance for a 45 degree diagonal edge is $\sqrt{2}$. Therefore, we use the method described in Eq. (2) and (3) to secure all the sheet structures having a thickness smaller than N_{set}. For example, we can securely detect 2-pixel thickness sheets positioned in vertical (2 pixels), horizontal (2 pixels), or diagonal ($\sqrt{2} \times 2$) with $N_{set} = 4$.

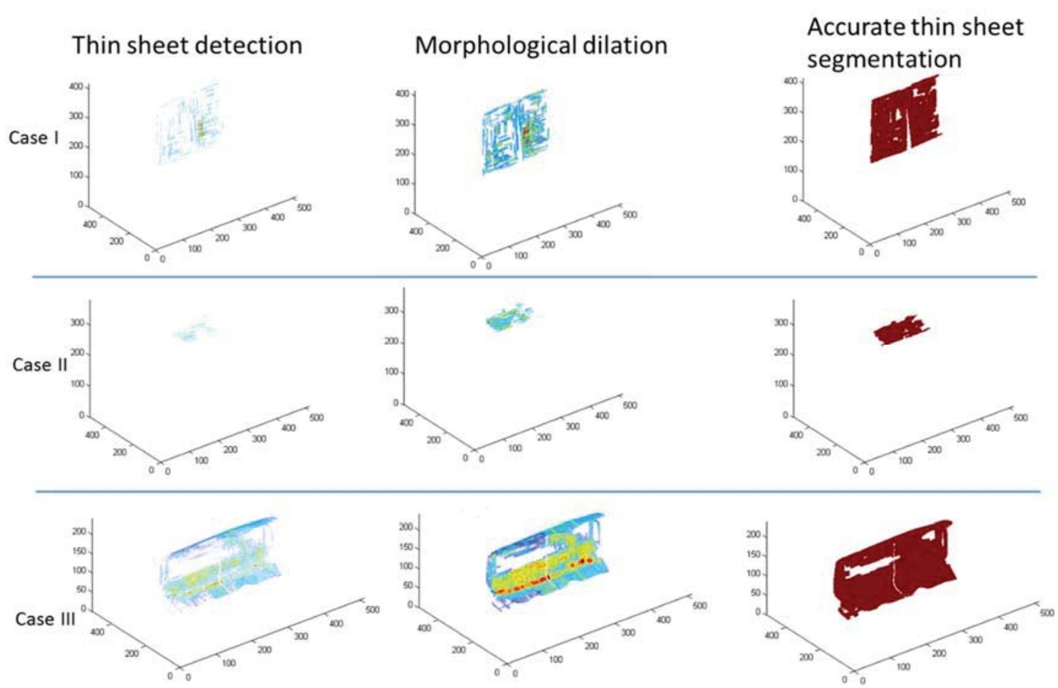


Figure 19. PT sheet detection results displayed in 3D space

Axial, coronal, and sagittal planes are processed to combine for the final results. The union of a collection of maps from axial, coronal, and sagittal planes was processed with a region property function, which removes isolated pixels. After the thin sheet detection, we apply morphological filter to segment PT targets accurately as show in Figure 19. The algorithm detects PT sheets accurately, because it considers the thickness of thin structures and can control the target thickness of sheet objects by changing a parameter in shift counter. (Figure 18)

5 Discussion and Conclusion

The presented SHPC technology is easy to adjust parameters for a few specific targets, which have high priority to detect. If we could generate an accurate look-up table and frequently update the look-up table, we don't need to re-train the whole algorithm. Moreover, it is possible to tune the parameters for an individual target performance without changing its overall system performance because we use a non-statistic table look-up method.

In the medical field we primarily use dual energy data sets or spectral imaging data sets (i.e., photon-counting detector data) for material segmentation and quantitative measurement of a specific target disease [12-14]. The SHPC algorithm has a great potential to be extended to the multi-spectral ATR since it is built on the histogram peak detection method, which can be easily implemented for a 2D histogram or a multi-dimensional histogram. From each spectral measurement, we can define a target peak and a spectral bandwidth for each energy band. As shown in other fields, the multi-dimensional spectral data can improve target detectability significantly. These kinds of applications can be found in medical field [15-17] and SAR imaging field [18, 19].

Additionally, advanced image reconstruction methods [20] could be helpful to improve PD/PFA through suppressing artifacts and improving image quality and resolution [21].

ATR was a challenging task and still is, there were multiple approaches that were discussed and tested. Through this process in the future, we could not only improve current version of algorithms and theories but contribute to the academia and the security industries.

6 References

[1] I. M. Bechwati, C. R. Crawford, M. Hiraoglu, and S. Simanovsky, "Apparatus and method for detecting sheet objects in computed tomography data," ed: Google Patents, 2000.
[2] J. W. Eberhard and M.-L. Hsiao, "X-ray computed tomography (CT) system for detecting thin objects," ed: Google Patents, 1998.
[3] B. Bhanu, "Automatic target recognition: State of the art survey," *Aerospace and Electronic Systems, IEEE Transactions on,* pp. 364-379, 1986.
[4] Z. Ying, R. Naidu, and C. R. Crawford, "Dual energy computed tomography for explosive detection," *Journal of X-ray Science and Technology,* vol. 14, pp. 235-256, 2006.
[5] S. Chemouni, K. A. Clark, K. Mann, T. R. Petrov, T. J. Rayner, and S. Skatter, "Explosives detection system using computed tomography (CT) and quadrupole resonance (QR) sensors," ed: Google Patents, 2005.
[6] J. W. Eberhard and M.-L. Hsiao, "X-ray computed tomography (CT) system for detecting thin objects," ed: Google Patents, 1999.
[7] S. Simanovsky, I. M. Bechwati, M. Hiraoglu, and C. R. Crawford, "Apparatus and method for detecting concealed objects in computed tomography data," ed: Google Patents, 2001.
[8] J. P. Thiran and B. Macq, "Morphological feature extraction for the classification of digital images of cancerous tissues," *Biomedical Engineering, IEEE Transactions on,* vol. 43, pp. 1011-1020, 1996.

[9] G. M. te Brake, N. Karssemeijer, and J. H. Hendriks, "An automatic method to discriminate malignant masses from normal tissue in digital mammograms," *Physics in Medicine and Biology,* vol. 45, p. 2843, 2000.

[10] B. K. Szabo, P. Aspelin, and M. K. Wiberg, "Neural network approach to the segmentation and classification of dynamic magnetic resonance images of the breast: Comparison with empiric and quantitative kinetic parameters< sup> 1</sup>," *Academic radiology,* vol. 11, pp. 1344-1354, 2004.

[11] R. Bhargava, D. C. Fernandez, S. M. Hewitt, and I. W. Levin, "High throughput assessment of cells and tissues: Bayesian classification of spectral metrics from infrared vibrational spectroscopic imaging data," *Biochimica et Biophysica Acta (BBA)-Biomembranes,* vol. 1758, pp. 830-845, 2006.

[12] A. Primak, J. R. Giraldo, X. Liu, L. Yu, and C. McCollough, "Improved dual-energy material discrimination for dual-source CT by means of additional spectral filtration," *Medical physics,* vol. 36, pp. 1359-1369, 2009.

[13] T. R. Johnson, B. Krauss, M. Sedlmair, M. Grasruck, H. Bruder, D. Morhard*, et al.*, "Material differentiation by dual energy CT: initial experience," *European radiology,* vol. 17, pp. 1510-1517, 2007.

[14] A. N. Primak, J. G. Fletcher, T. J. Vrtiska, O. P. Dzyubak, J. C. Lieske, M. E. Jackson*, et al.*, "Noninvasive differentiation of uric acid versus non–uric acid kidney stones using dual-energy CT," *Academic radiology,* vol. 14, pp. 1441-1447, 2007.

[15] Z. Liang, J. R. MacFall, and D. P. Harrington, "Parameter estimation and tissue segmentation from multispectral MR images," *Medical Imaging, IEEE Transactions on,* vol. 13, pp. 441-449, 1994.

[16] W. E. Reddick, J. O. Glass, E. N. Cook, T. D. Elkin, and R. J. Deaton, "Automated segmentation and classification of multispectral magnetic resonance images of brain using artificial neural networks," *Medical Imaging, IEEE Transactions on,* vol. 16, pp. 911-918, 1997.

[17] E. Frey, X. Wang, Y. Du, K. Taguchi, J. Xu, and B. Tsui, "Investigation of the use of photon counting x-ray detectors with energy discrimination capability for material decomposition in micro-computed tomography," in *Medical Imaging*, 2007, pp. 65100A-65100A-11.

[18] X. Zhang, L. Jiao, F. Liu, L. Bo, and M. Gong, "Spectral clustering ensemble applied to SAR image segmentation," *Geoscience and Remote Sensing, IEEE Transactions on,* vol. 46, pp. 2126-2136, 2008.

[19] K. Ersahin, I. G. Cumming, and R. K. Ward, "Segmentation and classification of polarimetric SAR data using spectral graph partitioning," *Geoscience and Remote Sensing, IEEE Transactions on,* vol. 48, pp. 164-174, 2010.

[20] S. Do, W. C. Karl, Z. Liang, M. Kalra, T. J. Brady, and H. H. Pien, "A decomposition-based CT reconstruction formulation for reducing blooming artifacts," *Physics in medicine and biology,* vol. 56, p. 7109, 2011.

[21] S. Do, W. C. Karl, S. Singh, M. Kalra, T. Brady, E. Shin*, et al.*, "High Fidelity System Modeling for High Quality Image Reconstruction in Clinical CT," *PloS one,* vol. 9, p. e111625, 2014.

7 Appendix: Presentation Cases

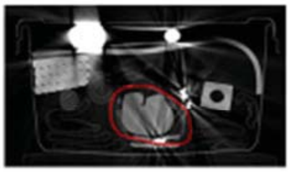
7.1 Presentation Case 1

Presentation Case (1) : SSN 13

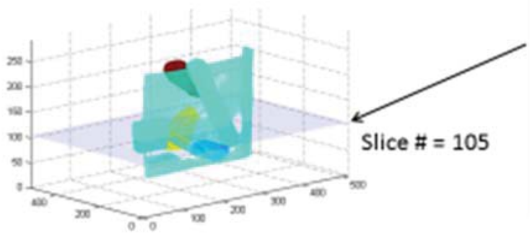
Case	Target	ID	Mass	SSN	Loc	Ort	Slice
Bulk with bad streaks cased by metal	Breast milk bag 10% Saline	6012	285	13	Bbb	z	105

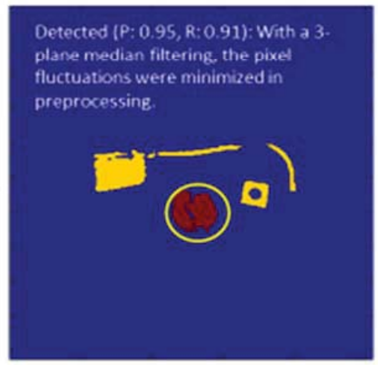
10% saline is segmented accurately even in the situation with streak artifacts.→ to make the target histogram sharp, we applied 3-plance median filtering, which mitigate the HU fluctuations on the target plane.This can be improved more with a warping based Metal Artifact Forecasting Method, which is shown in the backup slides.

CT image (10% Saline)



GT image





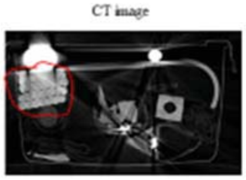
7.2 Presentation Case 2

Presentation Case (2) : SSN 13

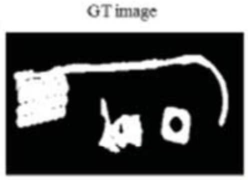
Case	Target	ID	Mass	SSN	Loc	Ort	Slice
Bulk with bad shading caused by beam hardening and scatter	Clay	6051	286	13	Abb	z	128

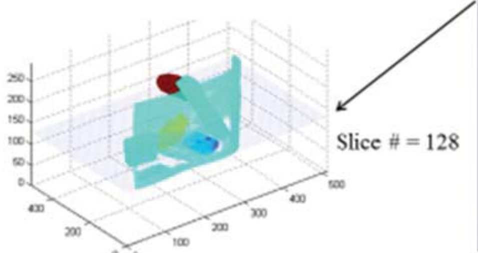
Clay was merged with a rubber sheet because a rubber sheet and clay was located too close and severe artifacts.
To detect target, we might need to apply stronger morphological filter with an additional MAR algorithm.

CT image



GT image



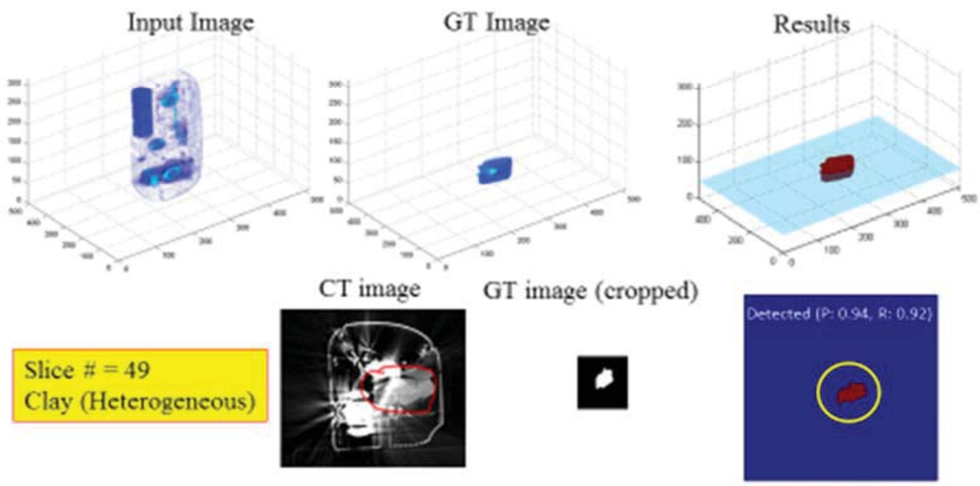


24

7.3 Presentation Case 3

Presentation Case (3) : SSN 35

Case	Target	ID	Mass	SSN	Loc	Ort	Slice
Bulk inside elect [illegible]	Clay	6150	290	35	Aab	z	49

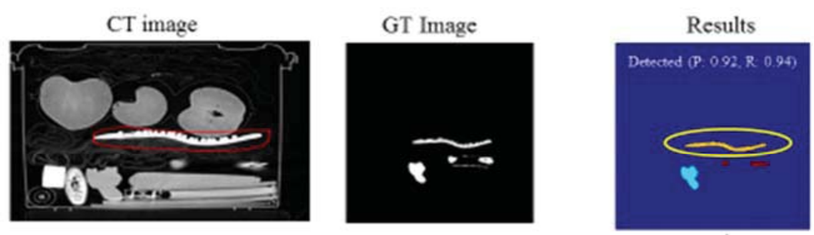


Even though there is strong metal artifacts, the core of clay target is detected accurately. The core extract in our algorithm is effective to find bulk targets.

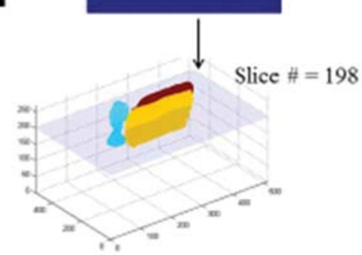
7.4 Presentation Case 4

Presentation Case (4) : SSN 193

Case	Target	ID	Mass	SSN	Loc	Ort	Slice
Bulk with texture	Clay w/ glass beads	6193	410	193	bbc	x	198



The bulk with texture target (i.e., clay with glass beads) is segmented accurately even though the material is not homogenous, and has an elongated shape.
→ The proposed algorithm could easily extract the core component, because the mass is big as in this case (Mass =410 cc) and the image quality is good (no artifacts).



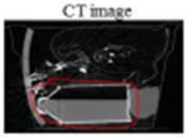
7.5 Presentation Case 5

Presentation Case (5) : SSN 63

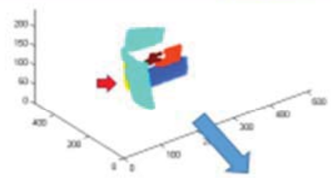
Case	Target	ID	Mass	SSN	Loc	Ort	Slice
Bulk with density close to water (~5% saline)	5% Saline – tin bottle	6163	274	63	bas	x	45

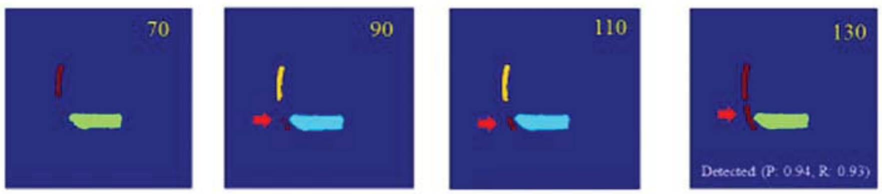
GT image

CT image



This case shows how target specific tuning can be executed for a rubber sheet without changing any parameters for the saline target.





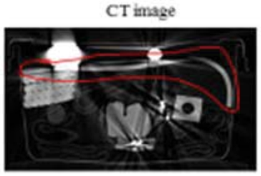
Target specific tuning: HUBW = 70, 90, 110, and 130 (Rubber sheet), Saline (no change)
The algorithm can detect homogenous bulk target effectively since the histogram is sharp and has no tails

7.6 Presentation Case 6

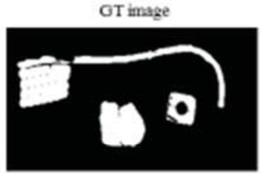
Presentation Cases (6) : SSN 13

Case	Target	ID	Mass	SSN	Loc	Ort	Slice
Sheet with bad streaks cased by metal, beam hardening and scatter	Rubber sheet 6 6mm	6018	685	13	Bcb	z	111

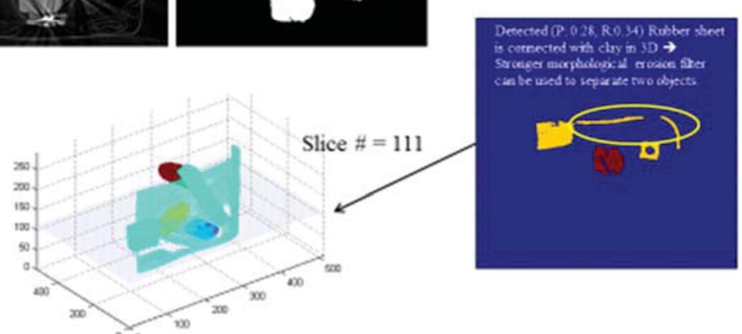
CT image



GT image



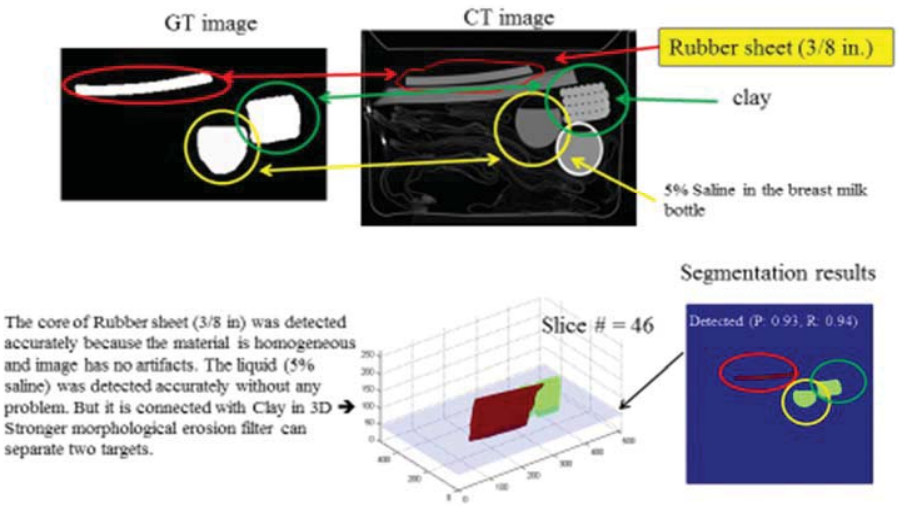
In the segmented results, rubber sheet is divided and combined to clay because of heavy artifacts. But the thresholds for recall and precision are 0.2 for both so it was detected.



7.7 Presentation Case 7

Presentation Case (7) : SSN 33

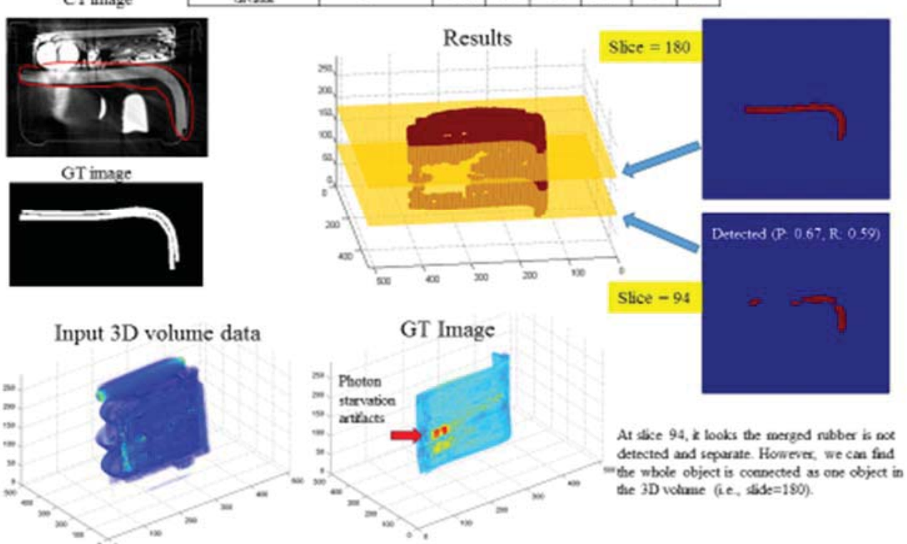
Case	Target	ID	Mass	SSN	Loc	Ort	Slice
Sheet laying on top of anot her flat object	3/8 rubber sheet on EE a magazine	6144	345	33	8ca	x	46



7.8 Presentation Case 8

Presentation Case (8) : SSN 11

Case	Target	ID	Mass	SSN	Loc	Ort	Slice
Object with lots of photon s tarvation	Merged rubber	7008	1360	11	8bb	z	94

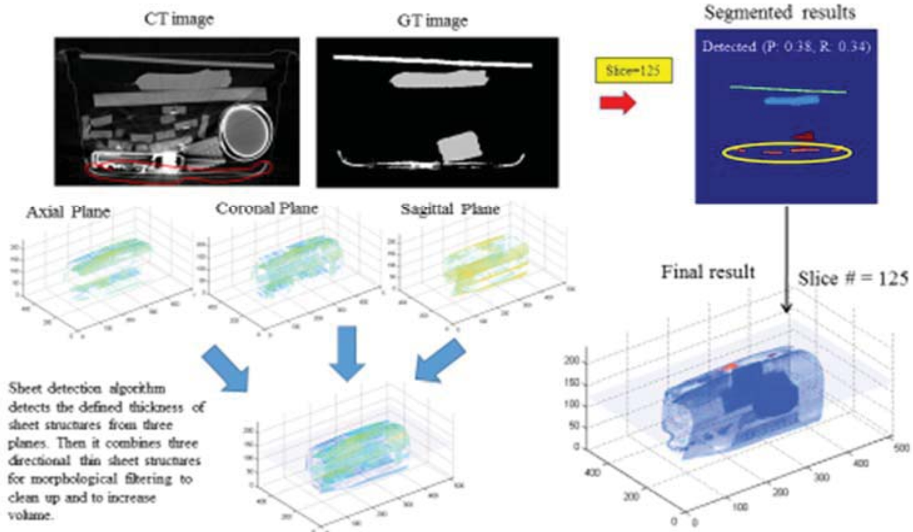


27

7.9 Presentation Case 9

Presentation Case (9) : SSN 18

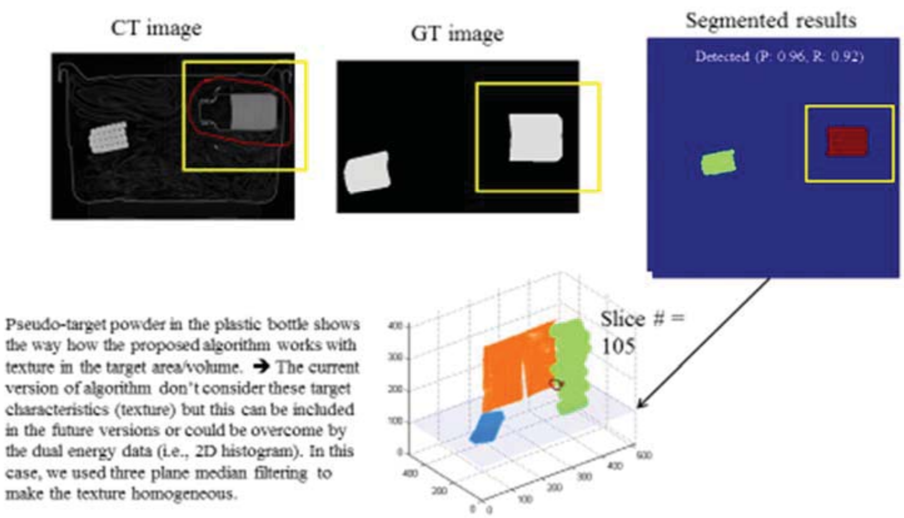
Case	Target	ID	Mass	SSN	Loc	Ort	Slice
PT sheet based on thickness	Neoprene rubber sheet 3.2mm	8026	350	18	Bab	Z	125



7.10 Presentation Case 10

Presentation Case (10/10)

Case	Target	ID	Mass	SSN	Loc	Ort	Slice
PT Powder (based on density, not mass)	TA_M800 plastic bottle+powder	6026	277	12	Cca	X	105



11.6.5 ATR Development – Top

"COE Task Order 4: LLNL Final Report on Automated Target Recognition"

COE Task Order 4: LLNL Final Report on Automated Target Recognition

Philip Top, Ana Paula Sales, Hyojin Kim, Stephen Azevedo, Harry Martz
Lawrence Livermore National Laboratory
Livermore, CA 94551

Work performed on the
Science & Technology Directorate of the
Department of Homeland Security
Statement of Work
HSHQPM-10-X-00045

Jan 23, 2015
LLNL-TR-665543

This work performed under the auspices of the U.S. Department of Energy by Lawrence Livermore National Laboratory under Contract DE-AC52-07NA27344.

COE Task Order 4: LLNL Final Report on Automated Target Recognition

Philip Top, Ana Paula Sales, Hyojin Kim, Stephen Azevedo, Harry Martz
Lawrence Livermore National Laboratory
Livermore, CA 94551

Executive Summary

Lawrence Livermore National Laboratory (LLNL) was tasked by the Department of Homeland Security (DHS) Science and Technology Directorate (S&T) with developing an automated target recognition algorithm for finding objects of interest in checked baggage at airports. The project used x-ray computed tomography (CT) data collected on a medical scanner of various simulated checked bags. Target objects of interest include clay, rubber, and saline in various concentrations. The final results are 93.9% probability of detection (P_D) with 11.9% probability of false alarm (P_{FA}) on all targets and 100% detection for pseudo target sheets. This result is achieved by using a three-stage algorithm made up of segmentation, post-processing, and feature extraction and object classification. The segmentation proved to be the most important stage and it involves merging two distinct segmentation results followed by a set of algorithms for cleaning up the merged segmentations. The first segmentation algorithm operates on planar 10x10 voxel slabs (two- dimensional segments) in the three spatial planes. The segmenter calculates a probability that each slab in the image is a part of a target, determines if the probability was above a threshold, and then merges all connected slabs together. The second segmentation algorithm is an ensemble segmenter, which generates a set of possible segmentations through random permutations and computes an average behavior. The distinct properties of these segmenters results in a merged segmentation with superior results than either segmenter alone. The merged segmentation is "cleaned up" using post processing techniques to achieve the final labeled images. The final classification stage made use of a random forest classifier algorithm to discriminate targets from non-targets based on a set of features of the segmented objects partially from the voxel slabs and partially from the individual voxels. Features include standard statistics such as mean, median, and standard deviations of the density, features capturing some aspects of texture, and some features about the surrounding bag structure. In addition to the detection and false alarm results above, we show how the algorithms can be trained to "overfit" to the limited data set provided and achieve near perfect results (P_D=100% and P_{FA}=0%). This over-training is not recommended for long-term operation, but instead is meant to demonstrate how test design is an important factor in aviation security. Possible improvements to the algorithm and the testing methodology (including blind testing) are also presented.

Table of Contents

List of Figures

List of Tables

List of Acronyms and Definitions

ALERT	Awareness and Localization of Explosives-Related Threats
ATR	Automatic Target recognition
AUC	Area under the curve
Bag	A package containing targets non-targets and pseudo targets
COE	Center of Excellence
CT	Computed Tomography
DCT	Discrete Cosine Transform
DHS	Department of Homeland Security
LLNL	Lawrence Livermore National Laboratory
Non-target	Something the ATR should not detect
PD	Probability of detection
PFA	Probability of false alarm
Precision	The fraction of a labeled detection which matches a real target
Pseudo-target	A target material with sub-minimum mass, sub-minimum thickness or a another material with density less than water
PT	Pseudo-target
Recall	The fraction of a real target which is captured by a labeled detection
ROC	Receiver Operating Characteristic curve
Segment	A labeled collection of voxels
Object	A segment which is intended to correspond to a physical target or non-target
S&T	Science and Technology Directorate of DHS
Target	Something the ATR must detect
TO4	Task Order 4
Voxel	A single volumetric pixel of a CT image

COE Task Order 4: LLNL Final Report on Automated Target Recognition

1 Introduction

The Department of Homeland Security (DHS) Science and Technology Directorate (S&T) has been supporting Centers of Excellence (COE) to encourage third-party participation in support of the DHS mission. The COE known as "Awareness and Localization of Explosives-Related Threats" or ALERT centered at Northeastern University is tasked with improving effective characterization, detection, mitigation and response to explosives-related threats facing the country and the world. A Task Order for ALERT (the fourth one, or TO4) was to develop automated target recognition (ATR) algorithms for finding target objects of interest in checked baggage at airports using computed tomography (CT) scanners.

Five academic and laboratory-based teams worked independently to develop these ATR algorithms. The Lawrence Livermore National Laboratory (LLNL) was one of the five participants and has developed an automated target recognition system that is described in this report. As a way to test the performance of each of the systems, ALERT acquired x-ray CT images of known targets in luggage bins that were used by all teams. The experimental data consisted of 3D single-energy x-ray CT volumes collected on an Imatron medical CT scanner. There were 93 different targets many of which were scanned in multiple bags, configured in luggage bins simulating bags[1] that were scanned and resulted in 188 volumetric CT datasets. The bags consisted of 421 target objects, 75 psuedo-targets and 1371 not target objects. An ALERT team acquired the data and identified the known "ground truth" by hand-labeling the objects in each image. ALERT also provided to the teams a common scoring program that accumulated the ATR output into a tabular format for cross-comparison. More details of the experimental methodology are available in (ALERT, 2014) .

For this project, the target objects of interest included clay, rubber, and saline in various concentrations, and a powder. The objects were placed in the bags in various shapes and configurations including blocks, bags, sheets, and other amorphous forms. The objective was to start with reconstructed voxel image data and create a set of labeled segments of the images where each label corresponds to a detection of an object of interest (target or non-target). According to the scoring criteria, a detection is counted if the label segment has a precision greater than 50% and a recall greater than 50% for bulk objects and a precision of 20% and recall of 20% for sheet objects. (Sheets, being more difficult to isolate for CT, were treated differently.) For pseudo-target sheets the precision and recall requirements were 10% for each. Precision is defined as the percentage of the labeled segment that overlaps with the labeled ground truth. Recall is the percentage of the labeled ground truth object of interest that is correctly labeled as an object of interest. The bags/bins were also packed with other objects such as water, electronics, other types of rubber, clothes and assorted other common objects. A full description of the test plan

[1] Since these bins were simulating luggage, we will refer to them as "bags" throughout.

is available in (ALERT, 2014). The available data consisted of 188 bags with various targets and labeled images with the ground truth. The ground truth was determined with a semi-manual process. All 3D reconstructed volumes and ground-truth labeled volumes were made available.

The final results with our system achieved 93.9% probability of detection with 11.9% false alarm with 100% detection for pseudo-target (PT) sheets. The full results are shown in Table 1.

Table 1. Final Results from Automatic object recognition at LLNL.

				No special rules (except for PT sheets)	
Object Type	Object Subtype	Level of Difficulty	Num Objects	Num Detected	PD [%]
Target	All	All	407	381	**93.6**
Target	Clay	All	111	107	96.4
Target	Rubber	All	158	150	94.9
Target	Saline	All	138	124	89.9
Target	Bulk	All	270	251	93
Target	Sheet	All	137	130	94.9
Target	All	Low	77	75	97.4
Target	Clay	Low	29	29	100
Target	Rubber	Low	22	22	100
Target	Saline	Low	26	24	92.3
Target	Bulk	Low	56	54	96.4
Target	Sheet	Low	21	21	100
Target	All	High	317	294	92.7
Target	Clay	High	82	78	95.1
Target	Rubber	High	125	118	94.4
Target	Saline	High	110	98	89.1
Target	Bulk	High	201	185	92
Target	Sheet	High	116	109	94
Pseudo-target	Sheet	High	10	10	**100**
			Num Non-targets	Num FAs	PFA [%]
			1371	163	**11.9**
				Num Scans with FAs	Avg Num FAs
				110	1.57

2 Algorithm Architecture and Philosophy

A general pipeline of the detection and classification algorithm is shown in Figure 1. Overall the system consists of three stages. The first stage consists of a pair of segmenters designed with different properties and a merging algorithm. The second stage consists of a collection of algorithms for cleaning up various issues with merged segmentations and specific algorithms for detecting and extracting sheets. The third and final stage contains a feature extracter for the labeled objects and a classifier to discriminate targets from non-targets.

A set of design principles guided the architecture. These included compatibility with new targets, and or different datasets; and separability or the ability to isolate and assess each individual algorithm components and replace it independent of the other components. Each stage of the process pipeline applies additional information extracted from the training data set to improve the final results and was examined and tested in isolation and in its relation to the entire process. Separability was maintained by defining an interface for each step, typically consisting of output files, and a testing framework that evaluated the effect of each individual algorithm on the results. Each stage of the algorithm could read in a file and output a file that could be later examined and tested or work in conjunction with the rest of the pipeline.

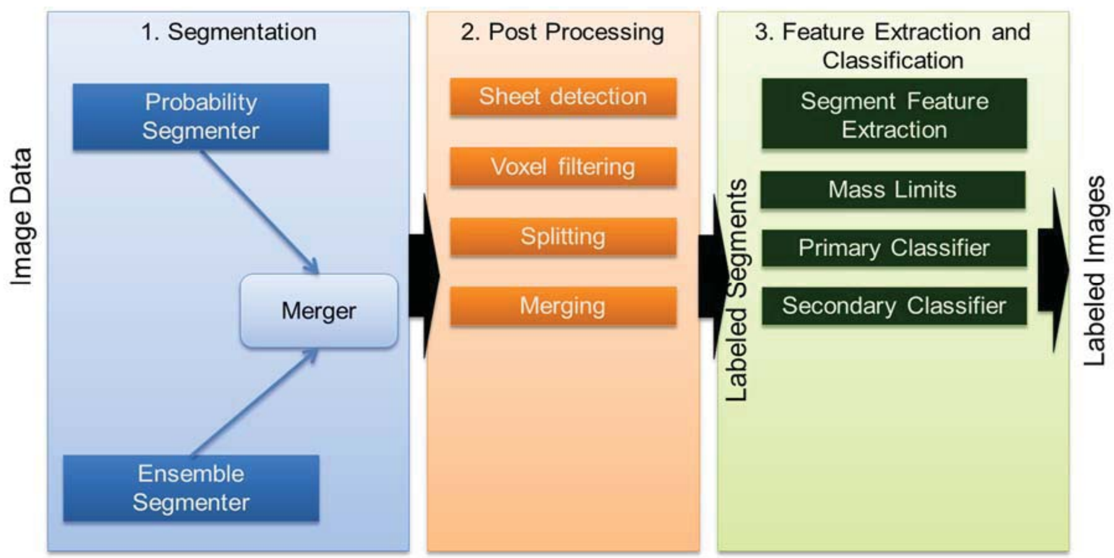


Figure 1. The ATR pipeline consisted of three stages.

3 Stage 1: Segmentation

The most complex part of the algorithm is the use of two independent segmenters. Early on in the development process it was determined that in order to meet the precision and recall specifications for this test the segmentation had to be as accurate as possible as future steps would rely on having accurate segments from which to extract features. It was also determined that the segmenter had to include some degree of classification in order to minimize false alarms. Based on the initial explorations, two distinct approaches were undertaken for the segmentation. The first, denoted herein as "probability segmenter", described in Section 3.1, was developed specifically for this project and the second, the "ensemble segmenter", was based on ongoing research into segmentation at LLNL that was developed prior to this project and tuned for the purposes of this project and is described in Section 3.2.

3.1 Probability Segmenter

The basic premise of the probability segmenter consists of labeling the voxels that have a high probability of being part of a target and removing all other voxels from further consideration. It operates in four steps:

1. Separate all the voxels into 10x10x1 voxel slabs and compute a set of features for each of them (further details are provided in Section 3.1.1);
2. Compute the probability that each voxel slab is a member of a specific target type based on the set of features (further details are provided in Section 3.1.2);
3. Apply a threshold on the probabilities in order to map them into binary values that indicate whether or not a voxel slab is a member of a target.
4. Connect all voxel slabs above a threshold into individual labeled segments. Voxel slabs with probabilities below the threshold are discarded.

The COE task established that in order for a labeled segment to be considered a detection, the segment had to capture at least 50% of the labeled ground truth voxels for bulk objects and 20% for sheet objects. Therefore, the aim of the probability segmenter is to identify at least 50% of the voxels from all bulk objects of interest as potential targets, and at least 20% for sheet targets. To improve operation of the segmenter, an additional internal goal for this algorithm was adopted which was to capture at least 70% of the voxels from 70% of the objects of interest. This second goal ensured that the more easily identifiable objects are segmented cleanly.

3.1.1 Voxel Feature Extraction

The image voxels were separated into 10x10x1 voxel slabs aligned with each of the three planes of the 3D image with 50% overlap. An illustration of these slabs in all three planes and the outlines of the different slabs are shown in Figure 2. Each slab consisted of 100 voxels, representing a tradeoff between statistical information and target resolution. Aligning along the three dimensions coupled with the small size of the slab allows detection of sheets in any alignment. In the case of the sheets, this approach is capable of capturing enough data of the thinnest target sheet in this test even at 45 degree angle with respect to the image alignment if the slabs are overlapped. This is based on the given dimensions of the minimum sheet size and the given voxel dimensions. The voxel dimensions are 1.5-by-1.5-by-1.5 mm. The minimum sheet width was 1/4 inch or 6 mm. At a 45 degree angle the sheet would occupy five voxels along one dimension, which covers half the voxels in some slab, allowing for statistics such as the median to be calculated based on these voxels. Some pseudo-targets had dimensions smaller than the listed dimensions in mass or thickness. The minimum target mass 250g.

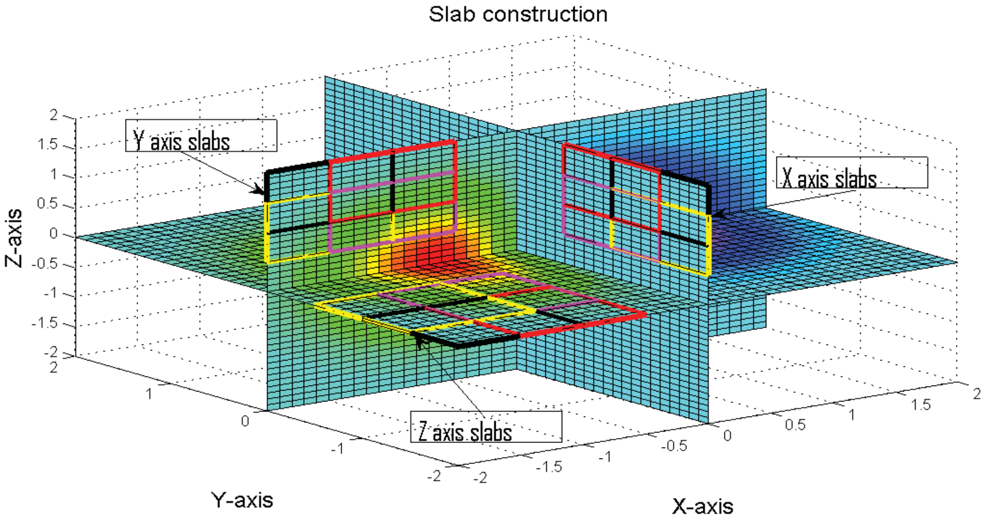


Figure 2. Voxel Slab Construction was performed in all three planes of the reconstructed CT volume.

For each of the slabs a set of 13 features was computed. The mean, median, mode and trimmean capture the density of the slab. The trimmean is the mean of the middle 90% of the data. The standard deviation, range, and 90% range capture the total variation in the slab. Six features based on the discrete cosine transform (DCT) of the data capture some aspects of the texture of the slab. Similarly to the Fourier transform, it uses sinusoidal functions as a basis for decomposing data. The Fourier transform uses complex exponentials, whereas the DCT uses only cosine functions and has the advantage of only producing real valued results. Taking the DCT of a 10x10 image produces a 10x10 result. For future discussion the element at position [i, k] will be referenced as dctik, so the element at position [2, 2] will be dct22. The dct11, or the result at position [1, 1], is equivalent to the mean value. The values along the diagonal capture the non-directional spatial variation in the data, which is primarily what we are interested in. Six features were extracted from the DCT: dct22+dct33, dct44+dct55, dct66+dct77, dct88+dct99, dct1010, and the sum(DCT)-trace(DCT). The trace of a matrix is the sum of the diagonal elements. The last feature captures all off-diagonal elements of the DCT. The diagonal elements represent rotationally invariant textural features. A detailed study was undertaken to identify the best classifiers and best features to use in discriminating between the different target types. The full details are available in Appendix A.

3.1.2 Voxel Classifier

In typical bag images, the majority of voxels are associated with non-target objects. The goal of the voxel classifier is to filter out the non-target voxels. The motivation for this step is two-fold. First, removing most of the non-relevant voxels should improve the accuracy of the segmenter, by reducing the number of decisions it needs to make. Second, by reducing the data to a small fraction of its original size, the complexity of the following steps in the segmenter, as well as in the subsequent steps of the ATR pipeline, are also reduced. Hence, an effective voxel classifier should make the entire ATR pipeline faster and more accurate.

The voxel classifier takes as input a vector of features for each voxel slab and produces as output the probability that the voxel slab belongs to a target object. By applying a threshold to the vector of probabilities for all voxel slabs, we obtain a binary value indicating whether or not each voxel slab belongs to a target object. The voxels that are classified as non-target are removed from the image and from any further processing. The remaining sparse image is then sent forward to the next step in the probability segmenter, the connected component algorithm. The goal of the voxel classifier is to remove as many non-target voxels as possible, while retaining the vast majority of the target voxels.

The voxel classifier consists of four distinct classifiers, one for each target subtype: clay, powder, rubber, and saline. Each of these classifiers takes as an input a vector of features of a voxel, and produces a binary output indicating whether the voxel is a target of the corresponding subtype or not. Voxels that are classified as a target by at least one of the subtype classifiers are kept in the image, and the voxels that are considered non-target by all classifiers are removed from the image. A diagram of the voxel classifier is shown in Figure 3.

The voxel classifier training is a supervised step, meaning that it learned the features associated with targets versus non-targets from labeled training data. Training data consisted of a small fraction(2%) of the voxel slabs associated with labeled target data. Inaccuracies in the ground truth can lead to errors in the training set but in this case the number of samples is large that the effect of the mislabeled voxel slabs is minimal. An exception to this is noted later for the case of some pseudo-target sheets. The training data itself consists of a set of training examples, in this case, a set of voxel slabs, each of which is represented by a label (target/non-target) and a vector of features. Once the voxel classifier is trained, it can then be used to classify new data. Training involves generating a set of decision trees using random choices for splits and only a small subsample of the training data each of which is setup to optimally classify the subset of data it is given. We explored a variety of supervised machine learning algorithms (see Appendix A), and obtained the best results using random forests (Breiman, 2001), which is the algorithm that we included in our ATR pipeline as the voxel classifier.

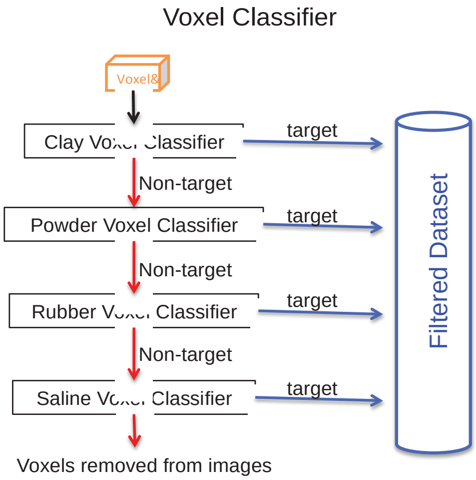


Figure 3: The voxel classifier is composed of four random forests, one for each target subtype: clay, powder, saline, and rubber. A voxel is classified as target if it is classified as positive by any one of the four random forests. Those voxels are kept in the images that are then subsequently sent to the segmentation step of the ATD pipeline. Voxels classified as non-target by the voxel classifier are removed from the images.

3.1.2.1 Random Forests

A random forest is a classification algorithm that consists of an ensemble of decision trees, where each tree is built with an element of randomness as described below. Each decision tree in the forest is a classifier that takes as input a feature vector and produces as output a binary classification, in this case, target or non-target. The random forest classifies a voxel slab by submitting its feature vector to each one of the trees and subsequently aggregating the outputs of all the trees into a final binary classification.

Decision trees operate by recursively partitioning the feature space of the data into exhaustive and mutually exclusive partitions. Each one of the partitions is based on a single feature and is associated with a label, in our case, target or non-target. Partitions are called “nodes” in the tree, and terminal nodes are called “leaves.” Given a training data set, a decision tree is built by recursively finding a combination of feature and threshold that best splits the data, where best is defined by how pure the derived partitions are in terms of labels. This process is greedy, that is, at each step the combination of feature and threshold that generates the best partition is chosen. This splitting is repeated recursively for each child partition until the partitions are pure (all training observations have the same label) or until a stopping condition is reached, such as a minimum leaf size Classification of a new observation is then obtained by interrogating its feature vector at each split until a leaf is reached. The new observation is then labeled based on the label associated with that leaf. Figure 4 shows a diagram of a simplified decision tree.

In random forests, each decision tree differs from all others to different degrees. This diversity is obtained by introducing randomness to the growing of the trees, as follows. Given a training set with N observations, each tree is grown using a training set consisting of N observations sampled at random and with replacement from the original training set. Given a feature vector of length *M*, at each split of each tree, a number $m << M$ of features is randomly selected out of the *M* features and the best split is chosen based on only the *m* features. Together, these two steps allow for diversity within the random forests, such that the ensemble is both more robust and effective than the individual parts. Inherently a single tree is highly over-trained and specific to a single training set, however, the random sampling of the training set and aggregation of multiple trees results in a much more robust classifier.

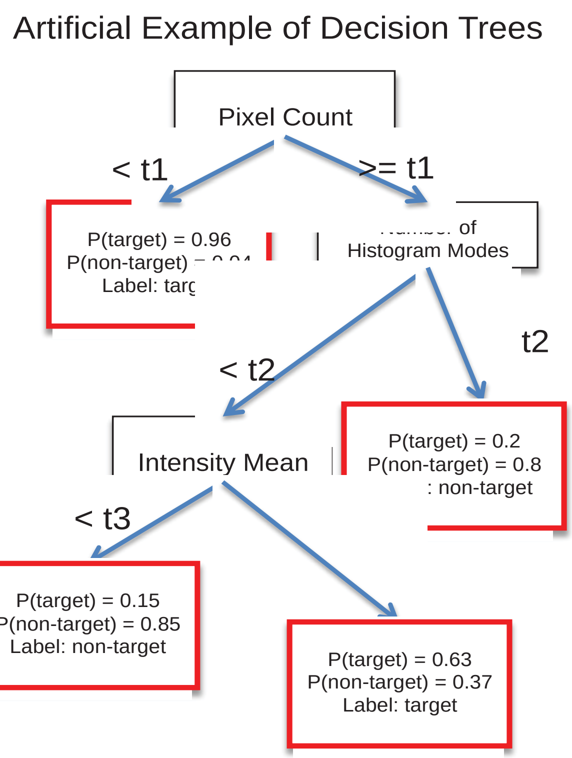


Figure 4: Decision Tree Example. Internal nodes are represented as boxes with black borders, and leaves as boxes with red borders. The labels within the internal nodes indicate the feature used at that node to split the data. Decision trees provide a partitioning of the feature space of the data into disjoint sets. Each partitioning is associated with a probability vector of the possible outcome classes. Classification of a new observation is obtained by mapping the features of the new observation to the partitioning of the data, until the new observation is associated with leaf. The class assigned to the observation is that associated with its leaf.

In addition to producing the best results, we chose to use random forests as the voxel classifier for a number of other reasons. Random forests are fast, they scale well both with number of observations as well as with the number of features. They are robust to anomalies. They account for interactions among features and non-linear relationships. Feature selection is obtained automatically, as part of the growing of the trees, and measures of relevance of each feature can be easily computed. Overall, random forests are very effective classifiers and we use them in our pipeline as the basis for a segmenter.

Another benefit of random forests is that there are relatively few parameters to tune compared to other classification algorithms. There are three primary parameters that affect the performance of random forests: the number of trees, the minimum size of terminal nodes, and the number *m* of features tested at each split. For each of these parameters there is a trade-off. Increasing the number of trees, up to a certain limit, increases the accuracy of the forest, but also increases the computational burden, both in terms of memory and computing power. Decreasing the minimum size of leaves allows for stronger individual trees at the detriment of computational performance. Finally, increasing *m* increases the strength of the trees (i.e., makes the individual trees better classifiers) but it also increases the correlation of the trees, making the ensemble of trees as a whole weaker. We have performed sensitivity analyses to understand how these parameters affect the performance of our random forests and to tune the forests used as our voxel classifiers. Some of the results of this exploration are described in Appendix A.

To summarize, the voxel classifier is composed of four random forests, one for each target subtype (powder, saline, rubber, and clay). Each random forest is composed of 500 trees, where the minimum size of the leaves is one. The number of variables tested at each split, *m*, is three. The details and rational behind each of these parameter choices is discussed in Appendix A. The voxel classifier labels as targets voxels that are classified as positive by at least one of the random forests. Voxels classified as target by the voxel classifier are kept, and those classified as non-targets are removed from the image. The sparse images containing only voxels classified as targets are then sent forward to the next step in the ATR pipeline, image segmentation.

3.1.2.2 Voxel Classifier Results

The results presented here were obtained using a voxel classifier, as described in previous section: four random forests (one for each subtype), each of which containing 500 trees, and having minimum leaf size of one and three features tested at each split. The complete set of features provided as input to the random forests consisted of the following summary statistics for each voxel slab: mean, median, standard deviation, range, dct22+33, dct44+55, dct66+77, dct88+99, dct1010, and dctA-trace. As described in Section 3.1.1, the features starting with prefix "dct" are elements of the inverse cosine transform. Overall, for all four subtypes classifiers, the mean and median of the voxel intensities were the most informative features (see Figure 5), followed by standard deviation and range. In the case of rubber the range is particularly useful likely due to the relative consistency of the density in the rubber objects Figure 10 shows the distribution of the range feature for the various object types.

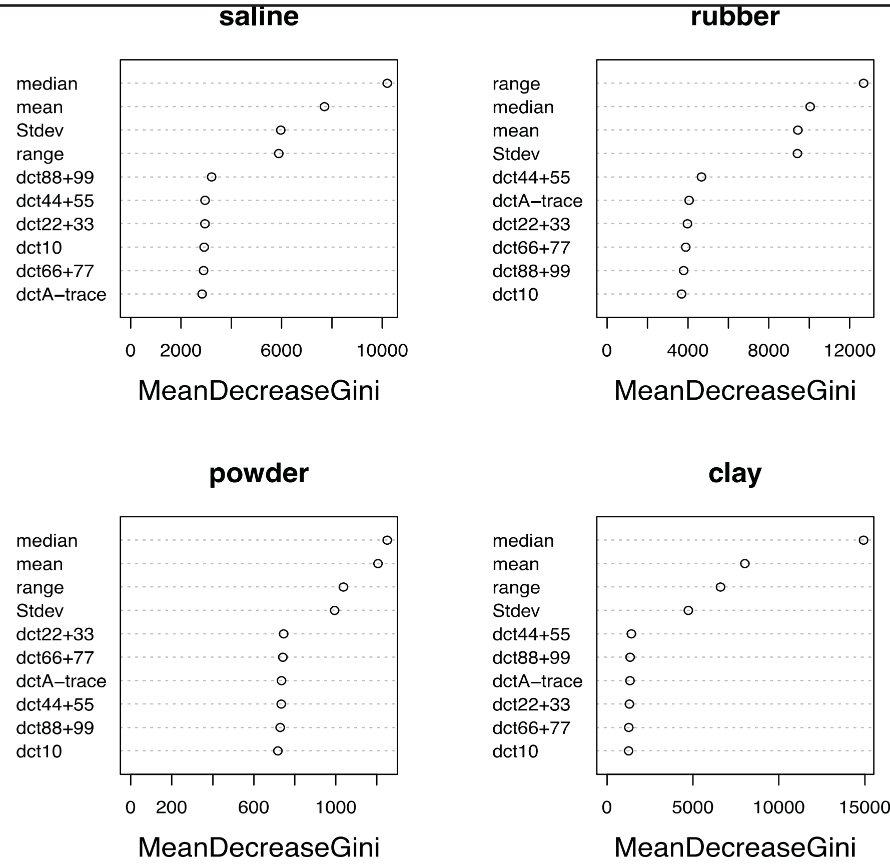


Figure 5: Relative importance of features for each of the four random forests (saline, rubber, powder, and clay) of the voxel classifier. The mean decreased Gini coefficient is a measure of how each feature contributes to the homogeneity of the nodes and leaves of the trees in the forest. The higher the mean decreased Gini coefficient of a feature, the more important it is for the random forest.

Table 2: Percentiles of number of voxel slabs of objects of each class that were correctly classified by the corresponding classifier.

	saline	rubber	powder	clay
0%	0.78	0.69	0.76	0.01
10%	0.82	0.72	0.82	0.65
25%	0.86	0.85	0.87	0.88
50%	0.92	0.91	0.89	0.93
75%	0.95	0.94	0.92	0.96
100%	0.98	0.98	0.98	0.99

In order to minimize over-training on the data, we perform three-fold cross-validation. In other words, the training data is randomly partitioned into three complementary sets, such that one of the sets is used to train the algorithm, which is then applied to evaluate the data in the remaining two sets. Using this scheme, voxel slab features are only evaluated by random forests that have not seen that observation while being trained.

All four subtype classifiers produce results with relatively high accuracy, with all of them having an area under the performance curve(AUC) at or above 0.9. AUC is the total area under a plot of the PD vs PFA created by varying the threshold at which a voxel slab is classified as target vs non-target and is a measure of the classifier performance. Overall, the best results were obtained for the clay classifiers (AUC ~ 0.97), and worst results for powder classifiers (AUC ~ 0.899), as shown in Figure 6. Table 2 shows the quantiles of the percentage of voxels of objects of a given class (saline, rubber, powder, and clay) that were correctly classified by the corresponding classifier. For saline, rubber, and powder, all objects had at least 69% of voxels correctly classified. The median of the fraction of voxels of objects correctly classified by the corresponding classifier was 0.92, 0.91, 0.89, and 0.93, for saline, rubber, powder, and clay, respectively.

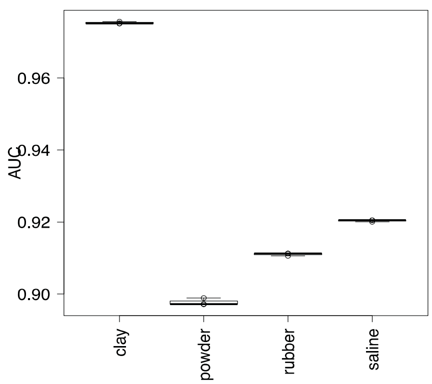


Figure 6 AUCs obtained by classifiers for the four object subtypes: saline, rubber, powder, and clay. Each box represents three AUCs, one for each of the three cross validation sets.

On a per-object basis, all target objects, except for one (object id = 7011), had at least 84% of its voxels included in the filtered set. Object 7011, a bulk object consisting of two blocks of clay merged, had 56% of its voxels included in the filtered dataset. The percentage of voxels of each object from the cleaned set retained in the filtered set is show in Figure 7. Objects 8024 and 8026 which were pseudo-target sheets were also characterized by lower detection rates likely due to the lower accuracy of the ground truth in those objects.

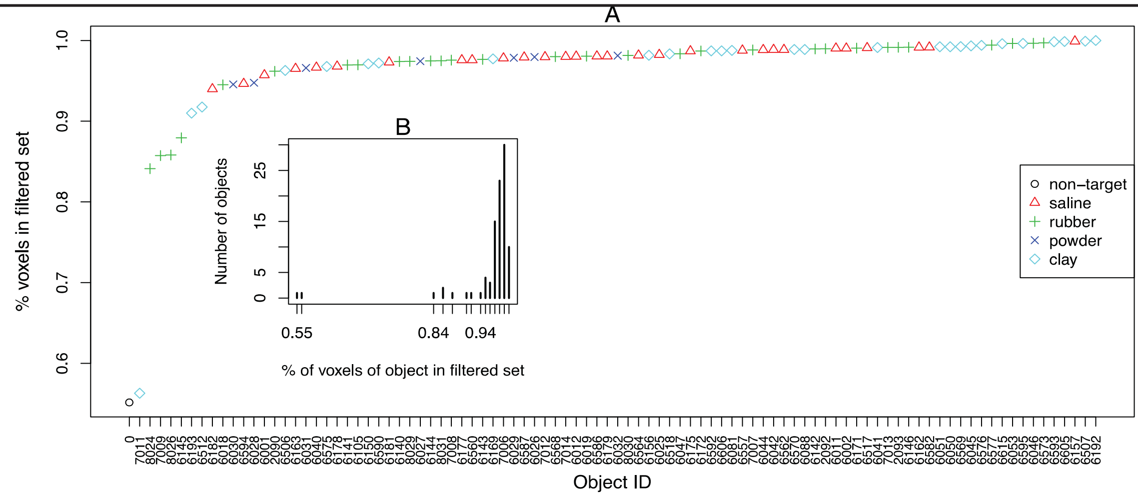


Figure 7: (A) Percentage of voxels of each object from the original dataset included in dataset filtered by the voxel classifier colored by object type. (B) Histogram of number of objects distribution as a function of the percentage of voxels included in filtered set.

3.1.2.3 Voxel Classifier Implementation

Using the information from the classification analysis we trained four classifiers using a subset of the calculated features. All four classifiers used the median, standard deviation, and range, and one of the DCT features. The rubber classifier used dct88+dct99, the saline and clay classifier used dct44+dct55 and the powder classifier which was later disabled since we were not required to detect powders, used dct22+dct33. Probability density functions for the various properties are shown in Figures 8-13. These figures illustrate that while the median is clearly the dominant feature for discrimination, the other features do provide some additional discrimination capability. Improved discrimination might be possible with ratios of DCT elements but this potential was not fully explored in the course of this project.

In order to better detect some pseudo target sheets and a few other anomalous objects it was necessary to select some specific bands in the rubber classifier and clay classifier to ensure sufficient detection of those objects. All voxel slabs with median intensity values of [1060 to 1150] for rubber and [2200 to 2300] for clay were labeled as high probability detections. The likely cause is due to some mislabeling in the pseudo target sheets along with a different type of rubber, and some deviations in the specific objects in question from the normal “clay-like” behavior.

Final results from the probability segmenter indicated that the high threshold captured at least 60% of the voxel slabs from 69% of the targets with a false alarm rate of 9%. A majority of these false alarm voxels are contained in a few objects or were removed, as they were isolated from other potential targets. The low threshold met the goal of detecting all targets at the required thresholds with a false alarm rate of 16% of the voxels in a bag. After the connected component segmentation numerous objects were merged together which reduced the PD significantly. A more sophisticated method of segmentation could have been used but it was not necessary when this method is used combination with the rest of the system.

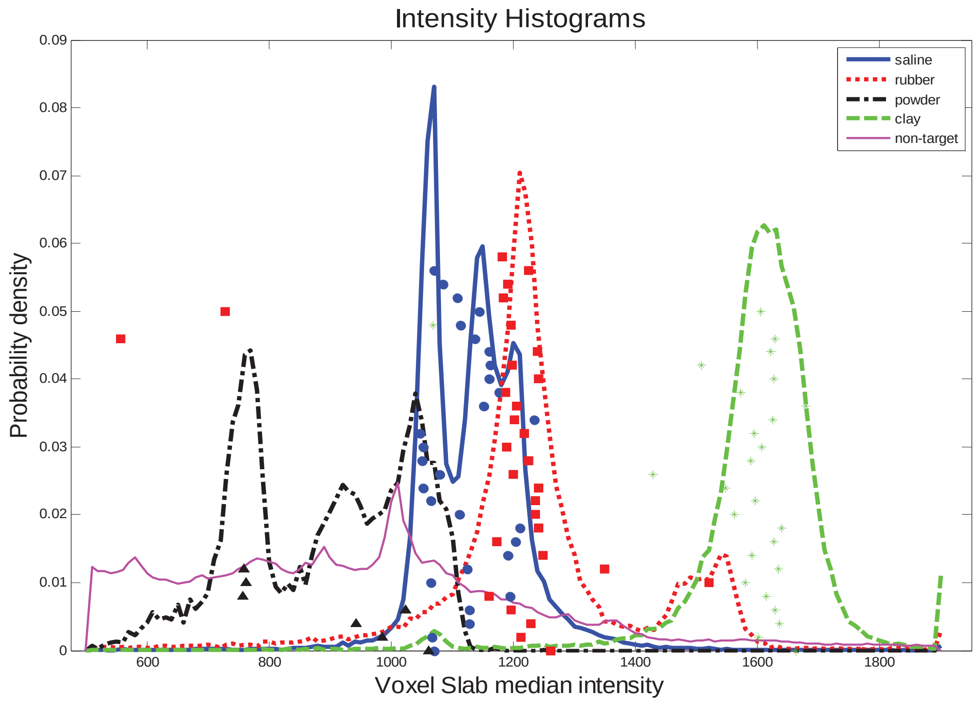


Figure 8 Voxel Slab Median Value Distributions

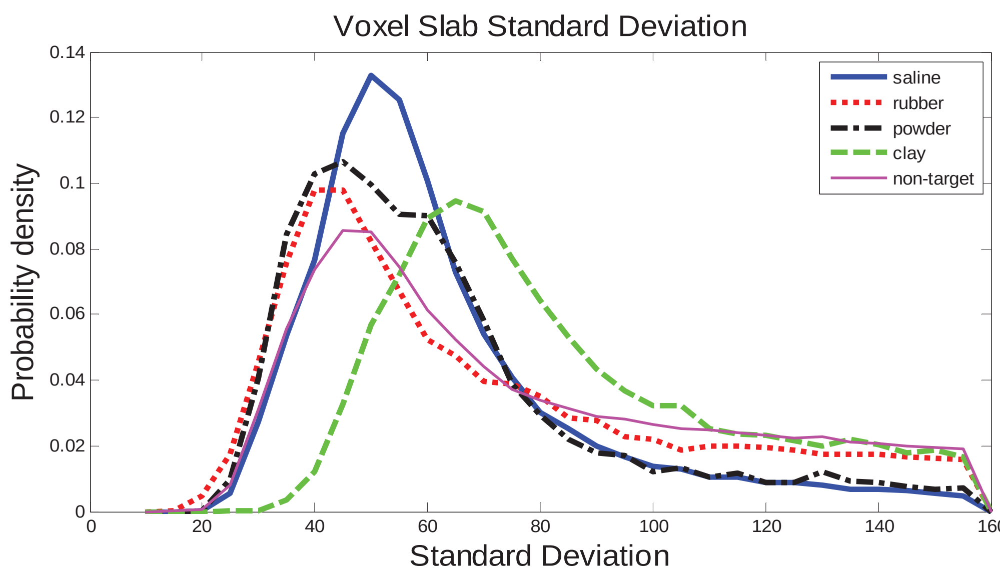


Figure 9 Voxel Slab Standard Deviation Distributions

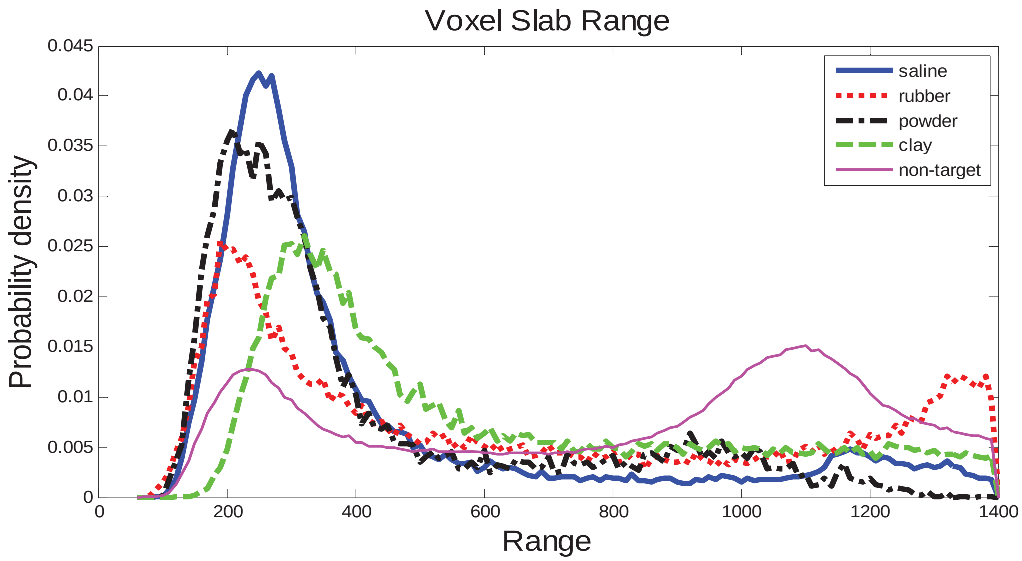


Figure 10 Voxel Slab Range Distributions

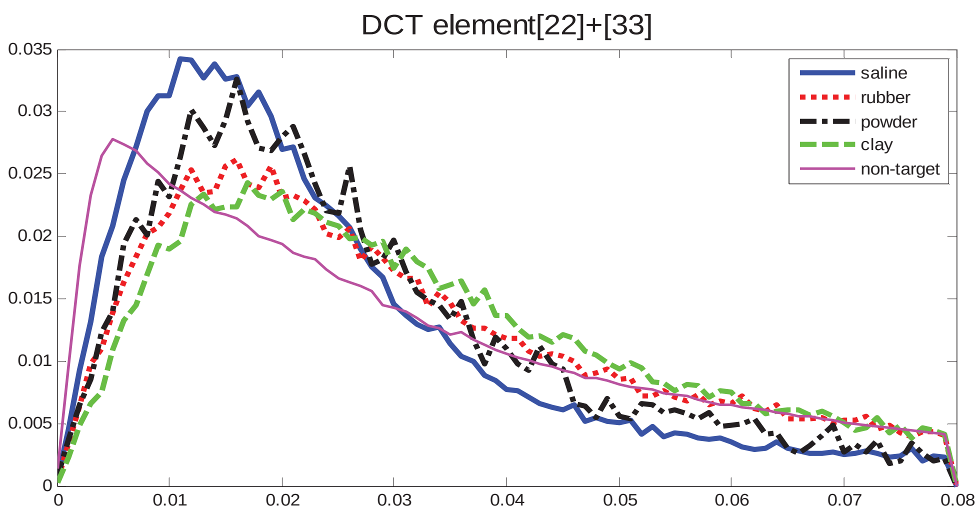


Figure 11 Voxel Slab dct22+dct33 distributions

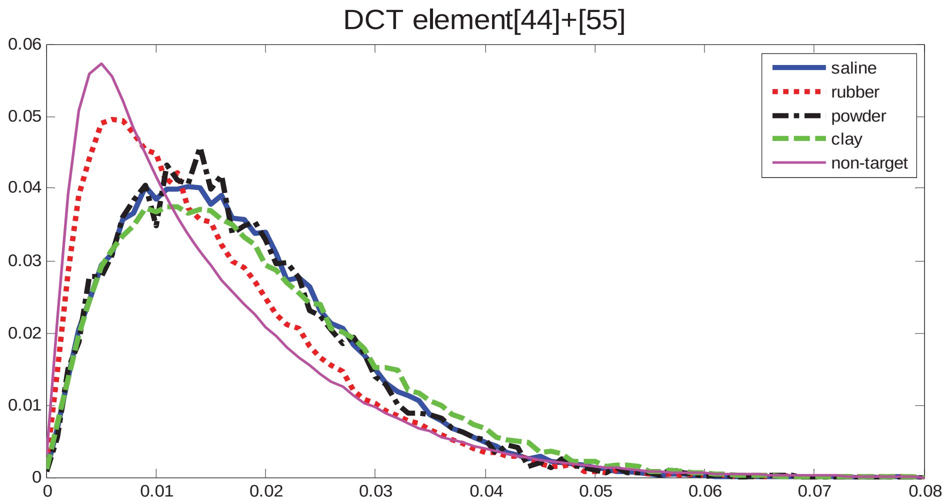


Figure 12 Voxel Slab dct44+dct55 distributions

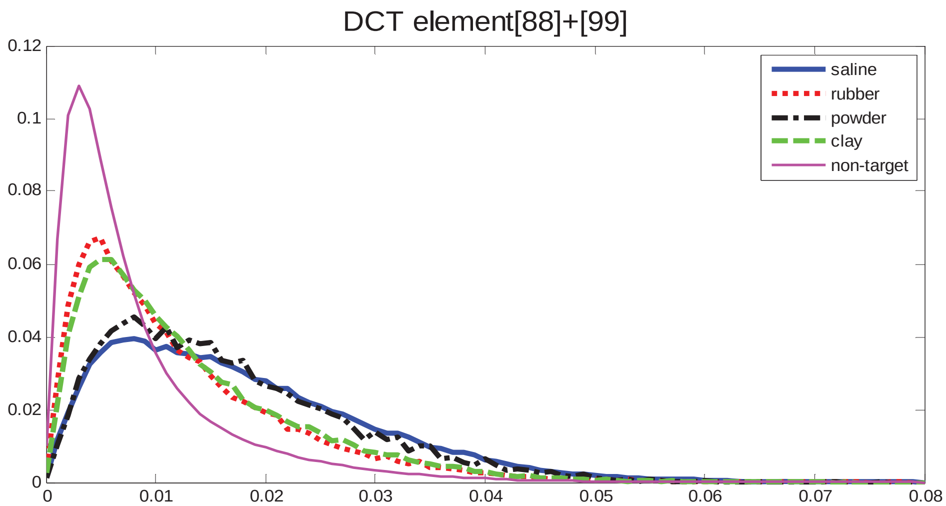


Figure 13 dct88+dct99 distributions

Each voxel slab was fed into each of the different classifiers and reported as a probability from 0 to 1 that the voxel slab contained target voxels. As part of the training, two thresholds were chosen for each target type. The higher threshold captured 80% of the voxels from 90% of the targets and the lower threshold captured 50% of all the targets based a small training set of data.

3.1.3 Connected Components Algorithm

The voxels identified as having probabilities above the high and low thresholds were merged using a connected components algorithm for each of the target types. The connected component algorithm identified all voxels as having neighbors in any of the three dimensions as being connected and part of the same object. Isolated voxel slabs typical of noisy pixels were removed from the labeled set. Single planes of voxel slabs were also removed from the connected object as this type of structure was typical of some forms of x-ray reconstruction artifacts namely metal streaking artifacts. The removal of these artifacts is an area that could benefit from additional analysis. For each target type areas with high probability were segmented and labeled before repeating the process with the low threshold areas. Segments that triggered thresholds for multiple target types were merged or split as appropriate. Segments formed from the higher threshold were kept separate while low threshold segments were merged with segments from other target types. This process produced a set of labeled segments for each image, resulting in a PD of 61% and a PFA of 91% of the final targets. The primary reason for the low positive detection and high false alarm rates was due to the fact that many targets were merged together by the simple segmentation process. Due to the constraints of the official test metrics (specifically that segments must meet the 50% recall and precision rates), merged targets can simultaneously reduce the positive detection and false alarm rates. Although the performance of the probability segmenter alone at this stage was underwhelming, when combined with the ensemble segmenter (see Section 3.2), it leads to improvements in both positive detection and false alarm rates, as discussed in Section 4. The purpose of this segmenter in the merged pipeline is to complement the ensemble segmenter. If it was operated alone additional complexity in the segmentation process and post-processing could improve the results significantly, but this was unnecessary for this test when merged.

3.2 Ensemble Segmenter

*
Note that this segmentation work was performed under the support of LLNL internal R&D funding (Laboratory Directed Research and Development Program) for "Coupled Segmentation of Industrial CT Images." The developed segmentation approach was applied to the problem of airport security and more specifically to TO4 data. We submitted a paper about the algorithm and experiments using TO4, which is currently under review and this section was excerpted from the paper (Kim, Thiagarajan, & Bremer, (Under Review) A Randomized Ensemble Approach to Industrial CT Segmentation, 2015):

We developed a highly flexible segmentation approach that uses an ensemble of hierarchical segmentations and exploits high-level semantic information to effectively find objects. This approach builds multiple hierarchical segmentations to explore as many potential ways of segmenting as possible, by randomizing the merging order of segment region pairs. Among potential segments in the ensemble, the most likely candidate regions are filtered, by incorporating high-level features defining objects. Finally, all localized candidates are combined into a consensus segmentation to produce the final segments.

This segmentation ensemble approach has several advantages: First, the multiple randomized segmentations, some of which are expected to be accurate or sufficiently close, compensate for a large variety of noise and artifacts; second, the global search for likely objects allows segments at multiple levels of the hierarchy; third, using a simple reference-based scheme, we compare segments to the training data, and robustly identify a set of candidate segments likely to describe objects of interest; Lastly, sequentially localizing all candidates for a particular object/region, the per-object consensus segmentation performs graph-cut segmentation to obtain a final object region, without the need to directly estimate the number of objects in the candidates.

3.2.1 Algorithm Overview

The algorithm begins with supervoxel-based over-segmentation to partition the entire volume into small-sized atomic regions, referred to as supervoxels. Then we construct an ensemble of bottom-up hierarchical segmentations from the initial set of supervoxels. Each hierarchy incrementally merges regions from the previous level. The edge affinity is measured as the similarity between their intensity histograms. Suppose $w^l_{i,j}$ is the edge affinity between two regions r^l_i and r^l_j in hierarchy level l. We compute $w^l_{i,j}$ as $w^l_{i,j} = exp(-\sigma \chi^2(H(r^l_i), H(r^l_j)))$ where $H(r^l_i)$ is the intensity histogram of region r^l_i, and χ^2 measures the chi-square distance between two histograms, and σ is the parameter for the Gaussian radial basis function.

To generate multiple independent segmentations from the same set of supervoxels, we randomize the merging order of candidate edges in each hierarchy level. See the paper (Kim, Thiagarajan, & Bremer, Image Segmentation using Consensus from Hierarchical Segmentation Ensembles, 2014) for more details about how to construct multiple randomized hierarchies.

Among all potential segments from the ensemble, we extract candidate regions likely to be an object of interest. For each potential segment, we first compute its feature set that consists of intensity statistics, shape features, area, and volume-to-surface area ratio. We then extract a semantic descriptor from the feature using local discriminant embedding (LDE), a supervised graph-embedding approach, to compare with supervisory data to determine whether or not it is a good candidate.

The final step is to segment multiple objects, given the candidate regions extracted from the segmentation ensemble. We segment object regions sequentially in the order of their confidence measures until all available candidates are examined, in which we do not need prior knowledge of the number of segments. We first sort all the candidate regions in the decreasing order of their confidence. We then pick the first available candidate region with the highest confidence and collect other available candidates that have a high volume overlap ratio with the first region. The collected candidate regions are combined into a graph-cut segmentation to obtain a final consensus object region. These regions are excluded from the candidate regions, and we continue the per-object segmentation if there is still available candidate region. Figure 14 illustrates an overview of the proposed approach.

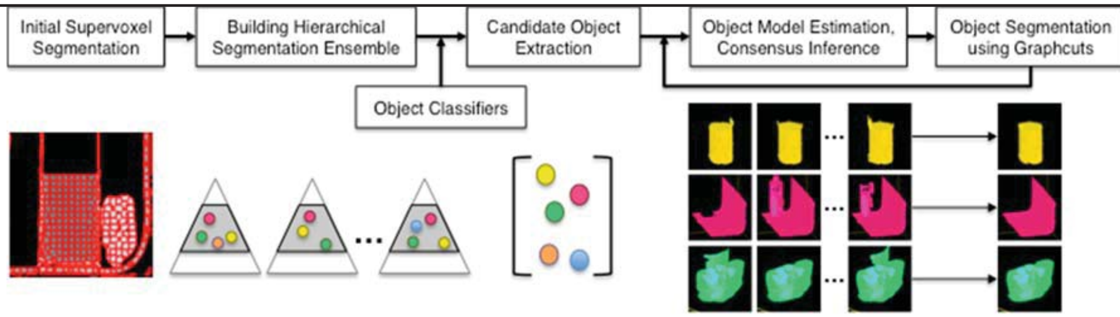
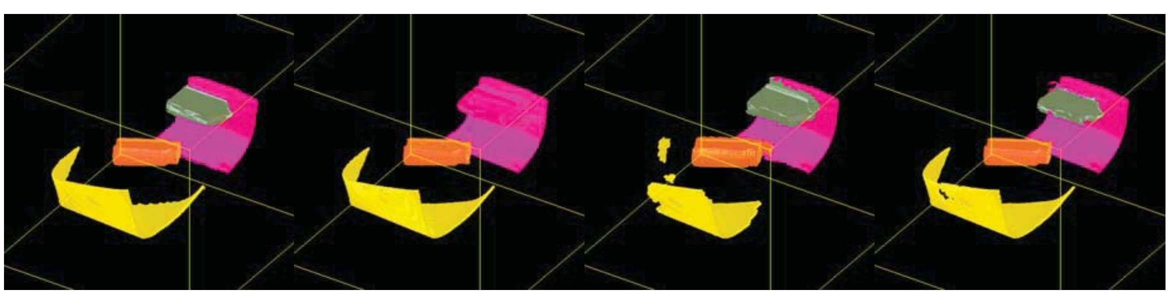


Figure 14 An overview of the ensemble segmenters

3.2.2 Experiments using TO4 Data

We applied our segmentation technique to the automatic target recognition (ATR) systems in the problem of airport security using the TO4 dataset. Here we concentrated only on the segmentation stage, aiming to provide a highly accurate segmentation of all target objects, without the target classification. For the quantitative evaluation of our segmentation, we trained our object classifier (an internal classifier within the ensemble segmenter system) using ground-truth labels of 4 different targets and pseudo targets. We extracted object features of all ground-truth regions and then extracted their LDE-based semantic descriptors. These semantic descriptors were used for extracting candidate object regions, as described previously.

In our setup, the only free parameter is the number of hierarchical segmentations, which was fixed to 20. We evaluated the segmentation performance of the proposed ensemble segmenter, in comparison to existing techniques: the popularly adopted region-growing technique and a semi-supervised graph-cuts approach, as shown in Figure 15. For the region growing, we manually tuned parameters for the TO4 dataset. For the semi-supervised graph-cuts, we simulated user's manual intervention by dilating the ground-truth label regions, and applied graph-cuts in each of those regions. Results from the scoring mechanism indicated and overall PD of 92% and a PFA of 45%.



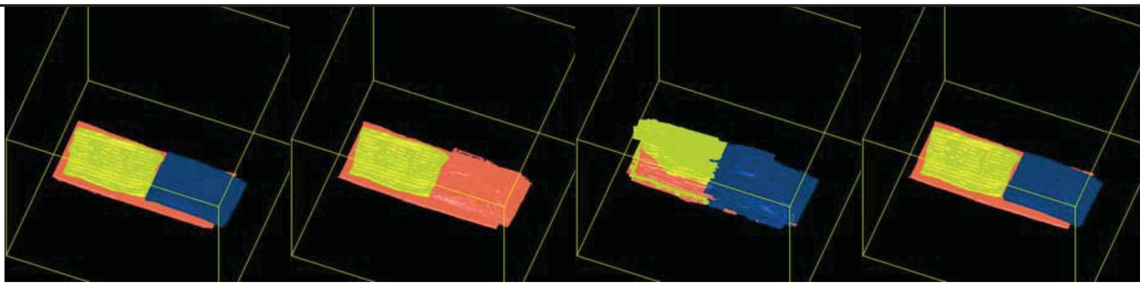


Figure 15 Segmentation results (SSN: 076, 148) of the proposed algorithm with other methods. From left to right in each row, the ground-truth labels, region growing, semi-supervised graph-cuts, and our ensemble segmenter

3.3 Segmentations Merger

The final step in the segmentation process is to merge the probability segmenter results with those from the ensemble segmenter. The process chosen takes the segmentation from the ensemble segmenter and adds new segments where the probability segmenter results contain labels not present in the ensemble segmentation result. This process resulted in a number of extra segments and small pieces but these were straightforward to remove of merge in the subsequent processing steps.

Merging segmenters in this way improves the overall results when combined with the post processing steps due to the different properties and characteristics of the different segmenters. The ensemble segmenter generally has better looking segments when graded by visual inspection but it can miss targets completely, particularly in the case of sheets. The missed targets represent a cap on the performance of the ensemble segmenter. Though the segments it does label are generally superior to the segments of the probability segmenter for the same object, the probability segmenter does not have the same theoretical performance limitations. Therefore, merging the two segmenters can raise the performance ceiling of the overall system. The downside of the probability segmenter is its tendency to merge many objects together, which is why the simple approach of basically using the ensemble segmenter results and adding any additional segments from the probability segmenter to the merged result tends to work. If it was desired to use the probability segmenter alone, additional complexity in the segmentation and post-processing steps would be needed to improve the results. Adding more complex merging algorithms into the merger step did not lead to any substantial improvement in the final result so they were not included in the final system.

4 Stage 2: Post Processing

Following the merge step was a series of post-processing algorithms designed to refine the segments. Numerous of examples of this type of processing are found in the existing literature on this subject. The patent from Telesecurity (Kwon, Lee, & Song, 2013)applies many similar techniques and further examples are found throughout the project report from TO3 (Crawford, Martz, & Pien). Sheet detection is also a topic of related patents (Eberhard & Meng-Ling, 1998). The distinguishing feature of the post-processing steps developed for this project is that they operate on voxel slab data vs. the original image data. This feature simplified the process significantly and allowed better characterization though at the expense of a small degree of resolution in the final segment. The voxel slabs were recalculated at the beginning of the process using solely voxel data from each label. The slabs started from the lowest indexed voxel from each label and progressed in steps of 10 for each dimension. Overlapping voxels was no longer necessary as experiments with different edges or overlapping blocks showed no change in the final results calculated by the scoring algorithms

Post processing consisted of a series of algorithms to split segments followed by a set of algorithms to merge segments. The first step is a sheet detection process. The thickness of the labeled segment is determined in each of the 3 dimensions. To be declared a sheet a segment had to have a thickness in one of the dimensions between 3 and 16 voxels for at least 25% of the segment area in that dimension and cover an area of at least 50 voxel slabs, or have thickness of between 12 and 32 voxels over 25% of its surface and cover an area of at least 80 voxel slabs. Once a sheet was detected all "clumps" on the surface were removed and declared a separate segment. This detection scheme was used to identify sheets and separate sheets from bulk objects with similar properties. A clump was identified as having a thickness greater than 2 times the sheet thickness and not be attributable to a sheet turning a corner.

The next stage is a statistical splitting routine. This algorithm computed the histogram of the median in the voxel slabs, searched for distinct peaks in the histogram and if the peaks met a set a thresholds the object was split. Specifically, the peaks had to be separated by at least 75 intensity levels, and the low point between them on the histogram had to be less than 85% of the peak counts. The split was then done at the minimum of a smoothed count value. The smoothing was done using a 20 point moving average filter. A check was then done to ensure that the voxel slabs from the different sides of the split were in fact in distinct locations in the image if so the object was split.

The third stage was a called a "cube split" The general idea is to count all the voxel slabs from a segment in a 10x10x10 cube then order these counts remove a certain fraction and do a connected components analysis on the remaining cubes. All large (>100 cubes) clusters of cubes were then split into new segments. The effect is to split objects that were close together but only touching on a thin set of points or were not touching to begin with as a result of the naïve connected components in the probability segmenter or merge process. This process achieves a similar effect to dilation and erosion which is another typical technique.

Following the split process, a final merge process was conducted. The primary statistic for merging is the mode of the histogram of medians of the voxels slabs of a segment. Distinctly labeled sheets with similar modes were checked for alignment and merged if they aligned. Then objects were checked for similar modes in the histograms, if the histograms matched and they overlapped, for large objects the shape and alignment was compared and if it matched they were merged, small objects were merged with less stringent criteria. The merging criteria were based on heuristics garnered from a subset of the data and subsequently applied to the entire data set. Based on the amount of overlap and the proximity of the histograms and the matching edges of the segment they were merged in entirety or partially. The exact threshold depends on the size of the segment and its mode.

Once the final segments were determined the voxel slabs were converted into a labeled image. The conversion process mostly consisted of mapping the voxel slabs back onto the voxel space in blocks of 10. However, all voxels with intensity values less than 475 or greater than 2300 were cut from the final labeled segment. Small objects that could not meet the minimum mass threshold established by this project were also removed. Mass per voxel was approximated by intensity.

5 Stage 3: Feature Extraction and Object Classification

The segmentation step labels segments as targets, but there are a large number of false positives. That is, a large number of the segments represent non-target objects. The goal of the "object classifier" is to filter out these false positive segments, reducing the probability of false alarm (PFA). Hence, the object classifier takes in a list of segments, and for each of those, it determines whether it should be labeled as target or non-target. In the best case scenario, the object classifier reduces the PFA of the pipeline close to zero, while maintaining intact the positive detection (PD) rate.
To complete this process a set of features was extracted from each of the segments.

1. Mean of original image voxels
2. Standard deviation of original image voxels
3. Mean of voxel slab medians
4. Median of voxel slab medians
5. Standard deviation of voxel slab medians
6. Median of voxel slab standard deviations
7. Median dct22+dct33
8. Median ((dct44+dct55)/(dct22+dct33))
9. Median voxel slab range
10. Mean Voxel slab range
11. Median dct44+dct55
12. Median dct88+dct99
13. Mode of voxel slab median
14. Mode of the original image voxels
15. Lower half width the mode – first bin with more voxels than 0.5*mode histogram count
16. Upper half width of the mode –last bin with more voxels than 0.5*mode histogram higher than the mode
17. (Optional) 2nd histogram mode peak if present
18. (Optional) 3rd histogram mode peak if present
19. (Optional) 4th histogram mode peak if present
20. Total Bag mean
21. Fraction of Bag with intensity >2300
22. Mean of segment bounding box+20 voxels in all dimensions

23. Fraction of bounding box with intensity >2300
24. fraction of voxels in segment between 1060-1150

The final feature was thought that it might help with identifying pseudo target sheets. A study similar to what was done for the probability segmenter was also done with the final stage classifier

5.1 Methodology

Similarly to the voxel classifier, for each observation the object classifier takes in a feature vector and produces as an output a binary classification: target or non-target. Whereas in the voxel classifier the input feature vector consisted of features extracted from individual voxel slabs, in the object classifier the features are obtained from the combination of all voxel slabs associated with a given segment. We used a total of 24 object features, as described above.

The object classifier consists of four random forests, each specializing in one of the four target subtypes: powder, clay, rubber, and saline. A segment is deemed to represent a target object if it is classified as a target by any one of the four classifiers. We avoid overtraining by using three-fold cross-validation. For more details on the methodology, see Section 3.1.2. The training set was built by using the provided ground truth to label the segments appropriately.

5.2 Results

One way to measure the performance of the object classifier is by comparing the performance of the labeled images produced by the ATR pipeline prior to the object classifiers against the performance of the same images filtered by the object classifier. Based on the output produced by the T04 evaluation tool, the PD for clay, rubber, and saline for the former set of images were respectively 98.2, 96.8, and 94.9, and the PFA was 43.1%. The object classifier was successful in drastically reducing the PFA, from 43.1% to 11.9%, while decreasing the PD by only very small amounts (PD for clay, rubber, and saline were respectively 93.6, 96.4, and 94.9%). Figure 16 shows the relative importance of the segment features in the classifiers. These results indicate that the mass limit is important in filtering out small segments and some measure of the object density is next. The rest of the features show some value but perhaps only in a limited number of cases. A number of examples of the detection results are shown in Appendix B along with some cases where the detection failed.

5.3 Improving Results for Corner Cases

Corner cases are situations where the correct classification requires an adaptation of the classification pipeline based on some specific property of the object being detected. The results described above were obtained by “honest” approach, where overtraining was avoided to the extent practical. This is necessary in order for the results to reflect expected performance of the ATR pipeline when applied to similar but not-yet-seen data. In order to further improve results and achieve the level of accuracy required by this competition, PD approaching 100% for all target subtypes and PFA close to 0%, we firmly believe that overfitting is a must, and that is what in fact the vast majority of vendors end up doing in order to meet the requirements to get certified. We show here how we can include additional steps, in the form of secondary classifiers that are put in place to capture the segments that were erroneously labeled by the classifier as false positives, and hence removed from the labeled images. These secondary classifiers are designed based on the specific instances misclassified by our pipeline, and hence when applied to the same set of images that were used to create them in the first place, can improve the performance of the pipeline to the levels of PD and PFA required. In fact, by modeling specific corner cases, we were able to get the PD for all low difficulty targets up to 100%, and reduce the PFA down to 1.1% using these data-derived rules. The problem is that these results are misleading, as we do not expect to see these rates of positive detections and false alarms in the future when the pipeline is applied to a new set of images.

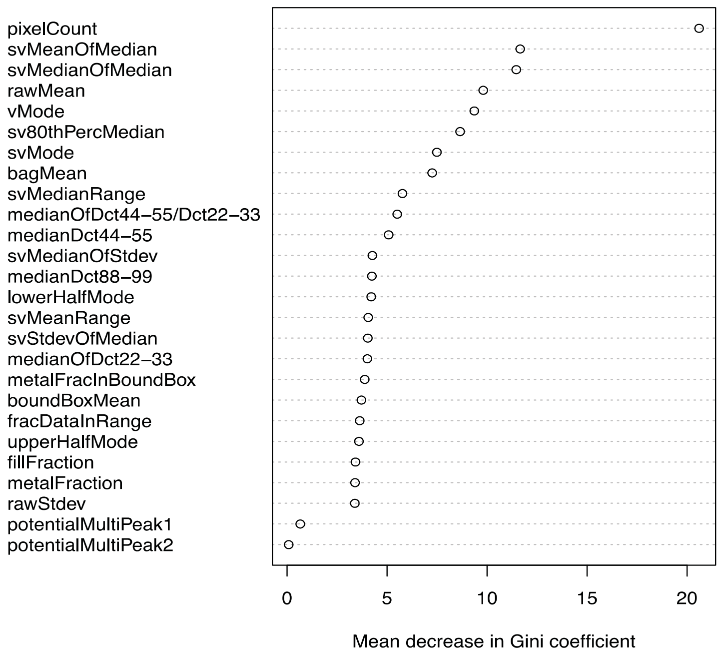


Figure 16: Relative importance of features in the construction of the object classifier. The mean decreased Gini coefficient is a measure of how each feature contributes to the homogeneity of the nodes and leaves of the trees in the forest. The higher the mean decreased Gini coefficient of a feature, the more important it is for the random forest[2].

---

[2] Abbreviations list. *pixelCount*: pixel count; *svMeanOfMedian*: mean of the median of contained supervoxels; *svMedianOfMedian*: median of the median of the contained supervoxels; *rawMean*: mean value of all single voxels; *vMode*: mode of the voxels; *sv80thPercMedian*: 80th percentile of the median of the contained supervoxels; *svMode*: mode of the supervoxels; *bagMean*: bag mean; *svMedianRange*: median range of the contained supervoxels; *medianOfDct44-55/Dct22-33*: median of dct44-55/dct22-33; *medianDct44-55*: median of dct 44-55 of contained supervoxels; svMedianOfStdev: median of standard deviation of contained supervoxels; *medianDct88-99*: median of dct 88-99; *lowerHalfMode*: lower half width mode (first bin with more voxels than 0.5*mode histogram count); *svMeanRange*: mean of range of contained supervoxels; *svStdevOfMedian*: standard deviation of median of contained supervoxels; *medianOfDct22-33*: median of dct 22-33; *metalFracInBoundBox*: metal fraction in bounding box;

6 Discussion

Of the missed targets, seven were due to inappropriate mergers with other objects, another 4 were due to inappropriate split targets, 1 object was split and merged with surrounding object and one object was corrupted by metal artifacts that prevented any segments from reaching the required detection threshold. The advantages of our pipeline are flexibility to new targets, data, and algorithms and ease by which new techniques and components can be integrated. Novel techniques include the use of texture based features in the probability segmenter and final stage classifier, and the use of multiple independent segmenters to improve the overall results, the use of voxel slabs for feature generation, and the use of random forest classifiers.

Both of the segmenters used contain characteristics that allow them to operate even in the presence of x-ray artifacts and thus no particular response to them was required. The probability segmenter masks over them with the use of voxel slabs and filtering of anomalous planes and the ensemble segmenter can overcome them with the help of the ensembles.

Object shapes play no role in the probability segmenter and only a minor role in the training of the ensemble segmenter. Shape was used in the extraction and detection of sheets but not elsewhere and none of the features used in the final stage classifier were based on shape. It is anticipated that similar targets in untested shapes would perform in a similar fashion to the current results.

Overtraining, except where intended, was minimized by sequestering a subset of the data, in the case of the ensemble segmenter; only using a small fraction of the data, in the case of the probability segmenter; and using three-fold cross validation for the object classifier. These methods do not entirely remove the possibility of overtraining—only limiting its influence to a certain degree. However, the nature of the project forced training to the test. The thresholds used in the probability segmenter and in some cases in the post processing methods are tied closely to the definitions of a detection and false alarm used in the experiment. In addition, many of the design decisions about where to put research effort are driven by the potential for impact on the final results.

6.1 Potential Improvements

Some initial explorations were done using complete image segmentation based on an available segmenter from Statovan (Woodhouse, June 2012). Stratovan was one of the teams working on TO3 for segmentation. The full report from this effort is available in (Crawford, Martz, & Pien). The results from the segmentation contained too many segments to be of practical use in this project. Furthermore, objects tended to be split

---

boundBoxMean: mean of the bounding box; *fracDataInRange*: fraction of the data between 1060 and 1150 (attempt to aid in pseudo sheet detection); *upperHalfMode*: upper half width mode (last bin with more voxels than 0.5*mode histogram higher than the mode); *fillFraction*: fill fraction of the bounding box; *metalFraction:* metal fraction in bounding box; *rawStdev*: standard deviation of all single voxels; *potentialMultiPeak1*: potential for single histogram peak; *potentialMultiPeak2*: potential for double histogram peaks.

into multiple segments requiring significant added complexity in the pipeline to merge them together. It is possible that with further assistance from the authors this segmenter could be better tuned to meet the requirements of this project, but for this project it was deemed impractical.

The selection of voxel slabs aligned with the 3 coordinate axes was sufficient for this project, however it is recognized that certain anomalous arrangements of targets would be problematic. There are a number of mechanisms for addressing this shortcoming. Additional planes could be formed with the 45 degree planes, this approach would likely be required if even thinner sheets were required to be detected, but it was not necessary to achieve the performance goals of this system. Sheets with thicknesses approaching the voxel dimension such as some of the pseudo target sheets are detected as long as a significant fraction of the sheet is aligned within approximately 20 degrees of one of the image axes, as they all were in this test. Though not necessary for this test further sheet specific algorithms are known and have been developed for similar applications (Eberhard & Meng-Ling, 1998). These algorithms could be applied instead of or in conjunction with the applied system.

The algorithms used in this project are all prototypes and could be refined significantly with further effort. A rigorous mathematical treatment of the probability segmenter with additional exploration of various features could lead to further reduction in false alarm rates. Likewise a rigorous treatment of the merger process and the addition of techniques previously developed in the COE task order on segmentation could also lead to further improvements. Some additional treatment of artifacts may be warranted particularly in the context of merging split object though these artifacts could also be dealt with by a better reconstruction. Most potential improvements rest in improving the segmentation results, which results in better features for the final classifier to act upon. Improvements in the overall system will be driven by improvements in segmentation as any appropriately applied modern classifier will tend to perform similarly with a given feature set. A significant potential area of improvement could rest in the refinement of the split and merge process. Cleaner segments could be achieved by further processing directly on the voxel data. Doing so was not necessary for the test criteria in place for this study.

6.2 Algorithm Discussion

As part of this project we conducted some research to determine what would be required to achieve 100% detection with 0% false alarm on the given data set as an example of how overtraining would play a role. Since there was no true blind testing which is really required to gauge true performance, the test results can be manipulated or skewed through overtraining both intentional and unintentional. While the ideal result was not achieved in practice, the achieved results come close and are shown in Table 10.

The achieved results reduced the PFA to 1.1% and increased the PD to 95.1. The distinction between the two results rests in the use of corner cases. Objects with specific properties can be treated differently than other objects throughout the detection process, as these clusters of objects become smaller and smaller at some point the final results becomes highly overtrained to the training set and the performance results from the test bear little resemblance to the actual performance of the system. To achieve the results shown small clusters (at least two objects) were formed based on various features of the segment such as mean intensity, standard deviation, and number and location of histogram modes to distinguish merging and splitting in the post processing steps for the final segmentation. And several secondary classifiers for different object types in the object classifier stage. If this were extended to single object clusters a cursory examination indicated another six objects could be detected improving the PD somewhat more. Applying similar techniques in the segmenter could potentially detect the remaining missed detections. It is clear that this result is over-trained but the boundary between overtraining and not is not particularly clear in this context since in some cases the clusters given special treatment could be for physical reasons in other cases they could be very specific to the test set and the only way to determine that is through the use of a series of truly blind tests.

Throughout development of the pipeline presented here many design decisions were made because they did not affect the final test results based on the test criteria given. The effect is that many of the details of the final system become tied to the exact nature of the test criteria. If those criteria were different the resulting pipeline would look very different. The design process by which it was developed would be substantially similar but the final pipeline would be different. For instance if the test metric were based on actual precision and recall scores instead of a binary decision, that would have necessitated the use of more operations at the individual voxel level to gain a few extra percentage points, whereas in present form such processing had no effect on the final score and were thus left out. Many design choices such as the complexity of the segmentation merge or the detail of the probability segmenter were made because based on the separability of the pipeline we determined that the return in the form of test results on effort on that part of the system was low compared with other parts of the system. Though more subtle than the blatant overtraining used to detect the corner cases this is definitely a form of tuning to the test, overcoming it would require a test scored by a several distinct metrics which capture different aspects of system performance. If future projects of similar nature are done significant thought must be placed on the exact nature of the scoring to ensure broad coverage and to prevent tuning to the test.

Table 3 Final Results with corner cases

				No special rules (except for PT sheets)		New rules added for corner cases	
Target Type	Target Subtype	Level of Difficulty	Num Targets	Num Detected	PD [%]	Num Detected	PD [%]
Target	All	All	407	381	**93.6**	387	**95.1**
Target	Clay	All	111	107	96.4	107	96.4
Target	Rubber	All	158	150	94.9	151	95.6
Target	Saline	All	138	124	89.9	129	93.5
Target	Bulk	All	270	251	93	256	94.8
Target	Sheet	All	137	130	94.9	131	95.6
Target	All	Low	77	75	97.4	77	100
Target	Clay	Low	29	29	100	29	100
Target	Rubber	Low	22	22	100	22	100
Target	Saline	Low	26	24	92.3	26	100
Target	Bulk	Low	56	54	96.4	56	100
Target	Sheet	Low	21	21	100	21	100
Target	All	High	317	294	92.7	298	94
Target	Clay	High	82	78	95.1	78	95.1
Target	Rubber	High	125	118	94.4	119	95.2
Target	Saline	High	110	98	89.1	101	91.8
Target	Bulk	High	201	185	92	188	93.5
Target	Sheet	High	116	109	94	110	94.8
Pseudo-target	Sheet	High	10	10	**100**	10	**100**
			Num Non-targets	Num FAs	PFA [%]	Num FAs	PFA [%]
			1371	163	**11.9**	15	**1.1**
				Num Scans with FAs	Avg Num FAs	Num Scans with FAs	Avg Num FAs
				110	1.57	15	1

7 Summary

In this project we have developed an automated target recognition system that can detect targets from baggage X-ray data. Specific techniques include the use of a probability based segmenter based on features from 2D voxel slabs, the use of multiple independent segmenters, and random forest classifiers, we used these techniques explore the system performance through various stages of operation and explore what would be required to achieve perfect results.

8 References

ALERT. (2009). *Algorithm Development for Security Applications October 2009 Workshop Final Report.* Northeastern University, Boston, MA.

ALERT. (2014). *ALERT ATR Project: Top-Level Technical Specifications Version 5.*

Breiman, L. (2001). Random Forests. *Machine Learning 45(1),* (pp. 5-32).

Crawford, C. R., Martz, H. E., & Pien, H. (n.d.). *Segmentation of Objects from.* Northeastern University, Boston, MA.

Eberhard, J. W., & Meng-Ling, H. (1998, Jan 27). *Patent No. 5712926.* US.

Hiraoglu, e. a. (2000, Jun 5). *Patent No. 6026179.* US.

Kim, H., Thiagarajan, J. J., & Bremer, T. (2014). Image Segmentation using Consensus from Hierarchical Segmentation Ensembles. *IEEE International Conference on Image Processing,* (pp. 3272-3276).

Kim, H., Thiagarajan, J. J., & Bremer, T. (2015). (Under Review) A Randomized Ensemble Approach to Industrial CT Segmentation. *Submitted to IEEE Computer Vision and Pattern Recognition (CVPR).*

Kwon, J., Lee, J., & Song, S. M. (2013, Jul 4). *Patent No. 2013/0170723 A1*. US.

Megherbi, N., Flitton, G., & Breckon, T. (Sept 2010). A classifier based approach for the detection of potential threats in CT based Baggage Screening. *Image Processing (ICIP), 2010 17th IEEE International Conference on,* (pp. 1833-1836).

Murphy, K. P. (2012). *Machine Learning: A Probabalistic Perspective (Adaptive Computation and Machine learning Series).* MIT Press.

Simanovsky, S., Hiraoglu, M., Bechwati, I. M., & Crawford, C. (2000, Feb 15). *Patent No. 6111974.* US.

Woodhouse, D. F. (June 2012). Automatic Segmentation of CT scans of Checked Baggage. *Proceedings of the 2nd International Meeting on Image Formation in X-ray CT*, (pp. 310-313). Salt Lake City.

9 Acknowledgments

The authors would like to thank Timo Bremer who leads an LLNL-funded R&D project on CT reconstruction and segmentation. The ensemble segmenter reported on in this report was developed under Dr. Bremer's project. This report was excerpted from a conference paper submitted to the IEEE Computer Vision and Pattern Recognition (CVPR) 2015, which is now under review. The authors would like to acknowledge the support and funding for this research by the Science & Technology Directorate of the Department of Homeland Security. This work was performed under the auspices of the U.S. Department of Energy by Lawrence Livermore National Laboratory under Contract DE-AC52-07NA27344.

Appendix A Voxel classifier Evaluation

A.1 Random Forest Parameter Tuning

Random forests have only a small number of parameters to be tuned. We have explored the impact of two of these parameters, leaf size and number of trees, in the performance of voxel classifiers based on random forests.

Contrary to decision trees, in random forests, trees are not pruned. They are grown until pure terminal nodes are reached or until a user-defined leaf size is reached. Here the size of a leaf refers to the number of training examples that are assigned to that leaf. We have compared the performance of random forests with minimal terminal node size of one, four, ten, and fifty observations, while keeping all other input parameters equal. As seen on Figure A 1, the random forests were very robust to variations in the size of the leaves, as this parameter had no significant impact on their performance.

Similarly to the analysis performed regarding the impact on the size of leaves on performance of the random forests, we compared the performance of forests with 10,50, 100, 500, and 1000 trees, while keeping all other input parameters equal. Figure A 2 shows the performance of the random forests of different sizes. The smallest forests, with ten trees, performed worst, showing a wide range of AUC. Overall, starting at forests with 50 trees or more, the performances of the classifiers, as measured by AUC, were equivalent to one another and stable. However, when we compare the impact of the number of trees per classifier type (clay, saline, rubber, or powder) we see that especially for powder, increasing the number of trees from ten up to 500 in the random forest leads to incremental improvements in performance. For this reason, we used forests with 500 trees for the results shown in Section 3.1.2.2.

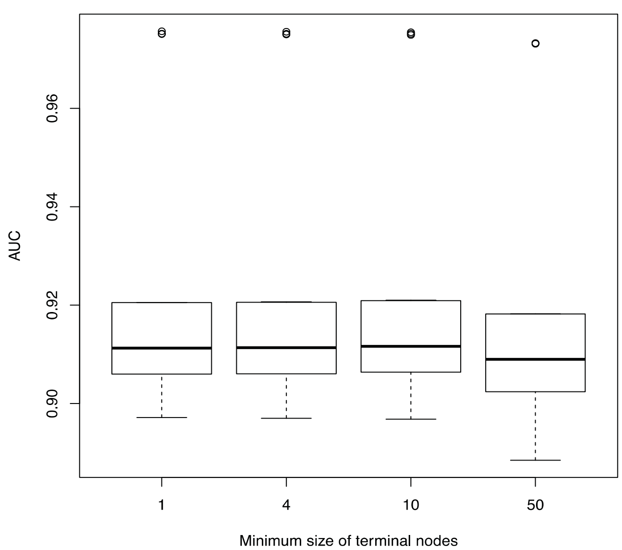


Figure A 1 Comparison of performance of random forest grown with different minimum size of terminal nodes

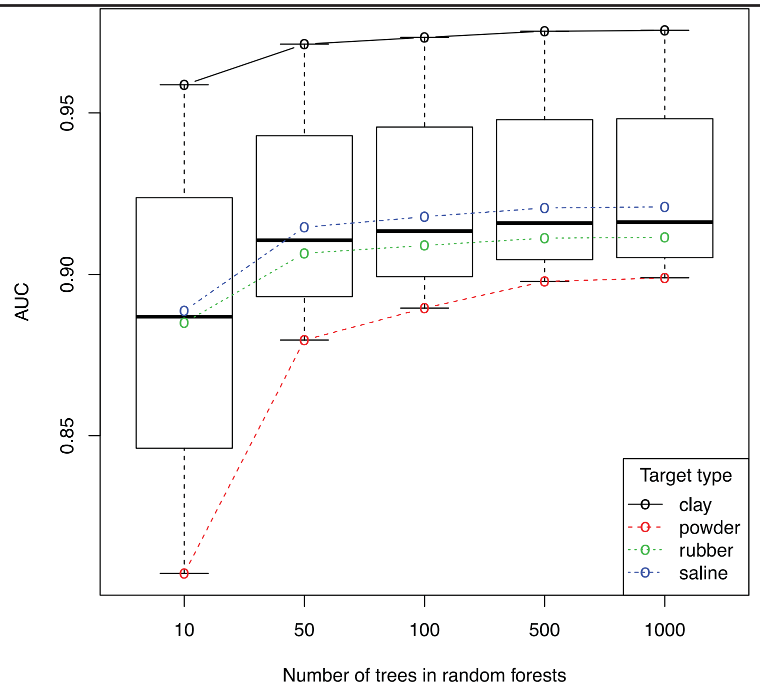


Figure A 2 Comparison of random forest performance based on number of trees.

A.2 Training Set Type

The bag data provided is unbalanced, that is, the number of voxels associated with non-targets is much larger than the number of voxels associated with targets. In particular, the percentage of voxels assigned to the categories non-target, saline, rubber, powder, and clay are respectively 79.1%, 5.78%, 9.12%, 0.93%, and 5.03%.

Training classifiers with unbalanced training sets can lead to biased classifiers in the sense that the majority class will be emphasized over the minority class. One approach to overcome this issue is to down-sample the majority class observations (in this case, non-target voxels) in the training data in order to obtain a balanced, or close-to-balanced, training set, which is a training set in which the number of positive and negative examples are approximately the same. We have built random forests using both balanced and unbalanced training set to assess how that affects the performance of the voxel classifiers.

For the unbalanced set, we used approximately 10% of the non-target voxels, but the training sets were still highly unbalanced. Table 3 shows the total number of positive and negative voxels in the training set for each type of classifier. For a classifier of a given type, only voxels of that type are considered positive voxels, and non-target voxels as well as the other three target types are treated as negative voxels.

The balanced training sets for each classifier were built to meet two constraints: 1) The number of positive and negative voxels must be approximately the same; and 2) The training set size must be as large as possible, being limited only by the number of positive voxels and the need to perform three-fold cross-validation. The negative observations of each type (non-targets, and the three remaining target types) were reduced via random sampling. This set containing all positive voxels and a sub-sample of the negative voxels was then subjected to random sampling to compose the three training sets for the three-fold cross-validation sets. Table A 1 shows the number of observations in the balanced training set of the four types of targets.

Figure A 3 shows the AUC of random forests built for each one of the four types of targets (saline, rubber, powder, and clay) when using balanced versus unbalanced training sets. Using balanced training sets consistently improved the AUC of the random forests.

Table A 1 Number of positive and negative voxel slabs in the unbalanced training sets for each of the four types of targets.

	Positive	**Negative**
Saline	28633	134039
Rubber	45439	117233
Powder	4584	158088
Clay	25073	137599

Table A 2 Number of positive and negative voxel slabs in the balanced training sets for each of the four types of targets.

	Positive	**Negative**
Saline	28778	26170
Rubber	45310	38383
Powder	4623	4516
Clay	24985	23610

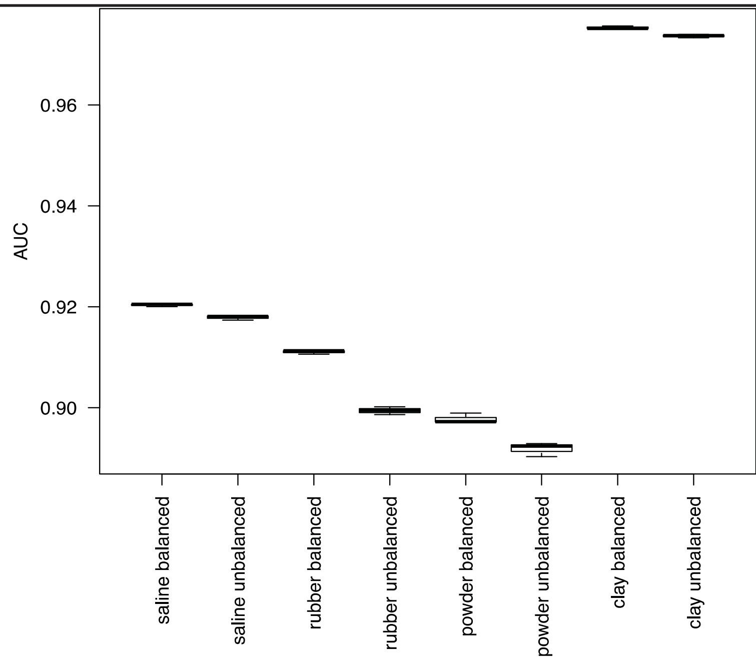


Figure A 3 Performance of random forests based on type of training set: balanced versus unbalanced.

A.3 Classification based on object ID

In an effort to achieve the best voxel classification results, we have explored the performance of random forests trained on a per-object basis. The rationale behind building classifiers per object is that this could potentially be an approximation of an upper bound of how well type classifiers could do, recognizing that any actual classifier designed in this fashion would likely be over-trained. The idea is that the variance in the voxels properties within each object should be no greater than the variance of the same voxel properties for voxels of a given type, regardless of the object ID. If this is the case, then object ID classifiers should yield better prediction accuracy than the type classifiers, whose results we have described in Section 3.1.2.1. Comparing the between and within object ID variance in some of the important predictor variables (such as voxel slab median) also suggests that building object ID classifiers may lead to improved results (see Table 5).

Hence, we have trained 93 different classifiers, one for each object ID (The number of object IDs of types saline, rubber, powder, and clay are respectively 29, 30, 7, and 27). Figure 11.A shows that there is a modest increase in the median performance of classifiers for all types, with the exception of clay. Interestingly, the variance in the performance of the classifiers has increased considerably in contrast to the performance of the type classifiers (compare Figures 6 and A 4.A). A large part of this increase in variance of performance is due to the variable size of each object. Overall, classifiers trained on larger objects (a.k.a., objects with a larger number of voxel slabs) tend to yield better performance accuracy, as measured by AUC. Figure A 4.B shows the dependency between classifier performance and number of positive instances in the training set (represented by the number of voxels of the corresponding object).

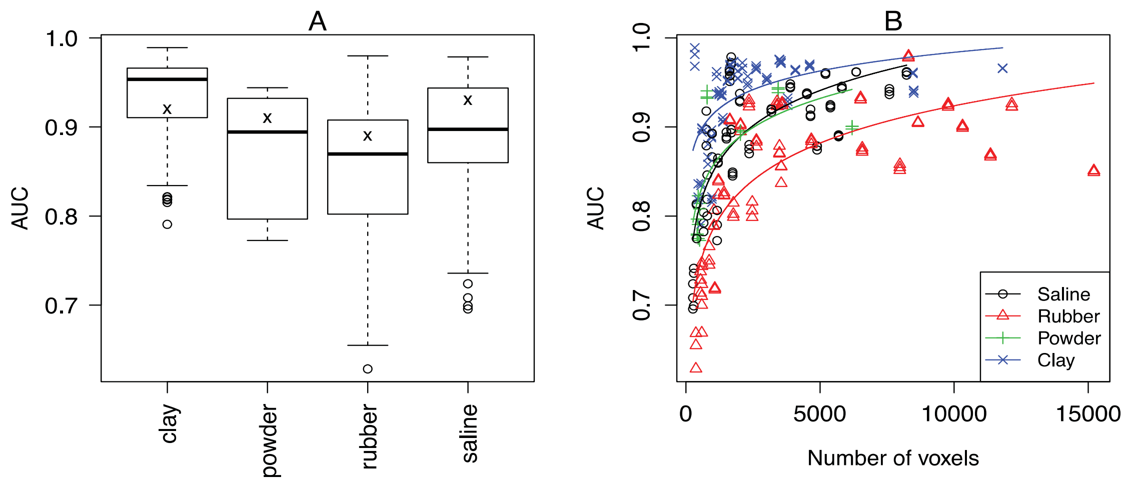


Figure A 4 (A) AUCs obtained by classifiers for the four object subtypes: saline, rubber, powder, and clay. Each box represents the results from three-fold cross-validation for each classifier of objects belonging to the corresponding target type (Number of classifiers of clay, powder, rubber, and saline were respectively 27, 7, 30, and 29). The "x" shows the median AUC of the three-fold cross-validation of the type classifiers shown in Figure 6. (B) AUC of the individual classifiers colored by the target type of each object ID. The solid lines show the regression of AUC as a function of the logarithm of the number of voxels. The R^2 of the regression for saline, rubber, powder, and clay were 0.61, 0.61, 0.53, and 0.31, respectively.

Table A 3 Anova model "Median ~ Type + ObjectId". Significant F values indicate that ObjectId contributes to the variance in the "Median" variable even after "Type" has been accounted for.

	Df	Sum Sq	Mean Sq	F value	Pr(>F)
Type	4	42722180843.98	10680545211.00	23067.82	0.0000
ObjectId	1	54542529.44	54542529.44	117.80	0.0000
Residuals	1492488	691031345655.29	463006.30		

A.4 Other Algorithms Considered for Voxel Classification

We considered a variety of machine learning algorithms as candidates for our voxel classifier, including random forests, support vector machines (SVM), adaptive boosting, nearest neighbors (NN), and naïve Bayes (NB)[3]. Overall, the best performances were obtained using variations of random forests, which was the primary reason for choosing this method for inclusion in our pipeline (further, secondary reasons for choosing random

[3] For details on these algorithms, see (Murphy, 2012).

forests over the other algorithms explored are described in Section 3.1.2.1). As explained above, the goal of the voxel classifier is to produce a filtered set that is a subset of the original data enriched for target voxels (that is, having as many as possible non-target voxels removes while retaining as many as possible target voxels). Table A 4 shows a direct comparison of the composition of filtered sets obtained by each of the five algorithms explored, both in terms of total percentage of voxels, as well as in terms of the percentage of target and non-target voxels. Naïve Bayes and SVMs tended to classify most voxels as "positive", only removing a small proportion of voxels from the filtered set, which was insufficient to adequately aid in the connect component segmentation. Nearest neighbors and random forests were the most successful approaches in reducing the size of the filtered set (both reducing the data to about 60% of the original number of voxels), with nearest neighbors actually being slightly more successful than random forests in this aspect. However, on a per-object basis, random forests were able to retain the largest fraction of positive voxels. Table A 5shows the percentage of voxels of each type (nothing, saline, rubber, powder, and clay) from the original dataset included in the filtered dataset.

Table A 4 : Summary statistics about filtered sets obtained via different algorithms. * value excludes object ID 7011, for which fewer voxels were detected.

	% total voxels	% non-target voxels	min % target voxels	min % voxels per obj
Random Forest	61.8	55	96	84*
Boosting	64.3	58	95	86*
Naive Bayes	92.2	91	98	86
SVM	87.4	85	97	90*
NN	60	53	91	80*

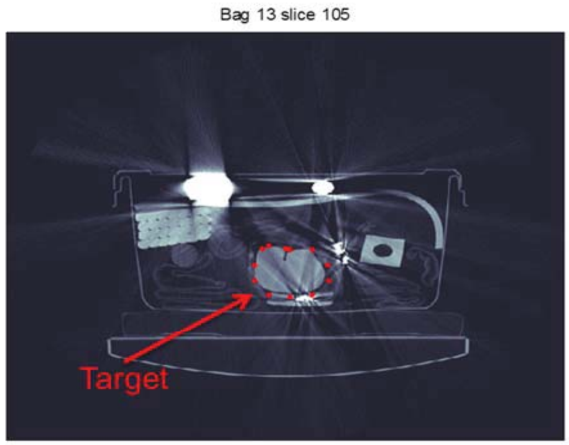
Table A 5 Percentage of voxels of each type (nothing, saline, rubber, powder, clay) from the cleaned dataset included in filtered dataset based on classifier type, using a threshold to obtain voxel TPR >= 0.9. NN_b = nearest neighbor algorithm using a balanced training set (i.e., a training set with equal size classes).

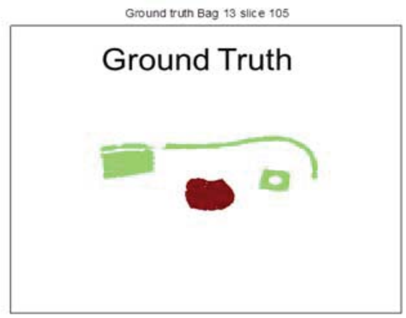
	nothing	saline	rubber	powder	clay
Random Forest	0.55	0.98	0.97	0.96	0.98
Boosting	0.58	0.97	0.97	0.95	0.98
Naive Bayes	0.91	0.98	0.99	0.99	0.99
SVM	0.85	0.99	0.99	0.97	0.99
NN	0.53	0.98	0.97	0.91	0.98
NN_b	0.63	0.98	0.98	0.97	0.99

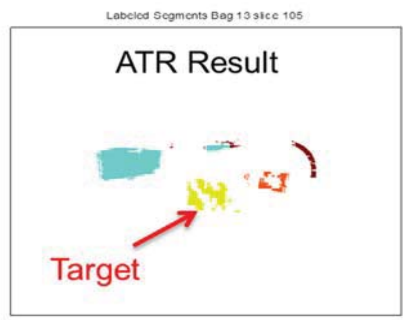
Appendix B Example Results

We show here a number of example cases of the results generated with the described pipeline along with some commentary on the results. All the cases show 3 images including a view of the image data. The same slice of the labeled ground truth data and another image of the labeled image data output from the ATR pipeline. The original image data is shown in grayscale with the white being the highest density. The upper limit on the image color scale is 2500. The target itself is surrounded by red dots. The ground truth data is shown with a white background the object in question is shown in red and other labeled targets are shown in green. The labeled image data is shown with a different color for each individual label. The colors appear in somewhat random fashion. For each of the cases examined, the actual precision and recall is shown along with a brief discussion about the situation and the effect on the results.

Case #1: Bulk with bad streaks caused by metal



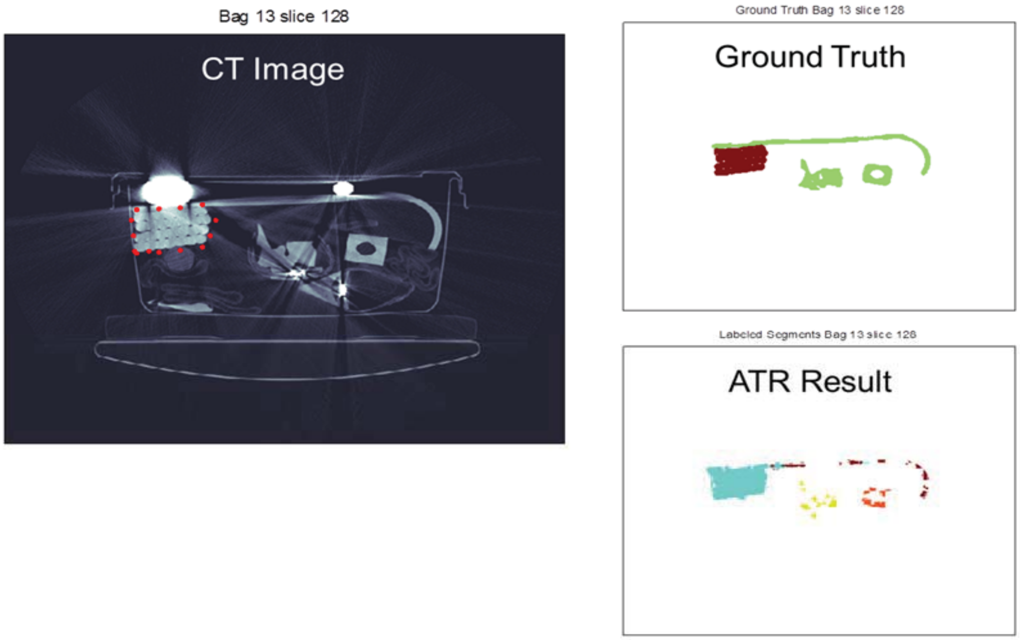




Detected: YES
Precision: 95.2%
Recall: 60.1%

Due to the streaking artifacts some parts of the target object were removed in the final image, thus the recall is reduced.

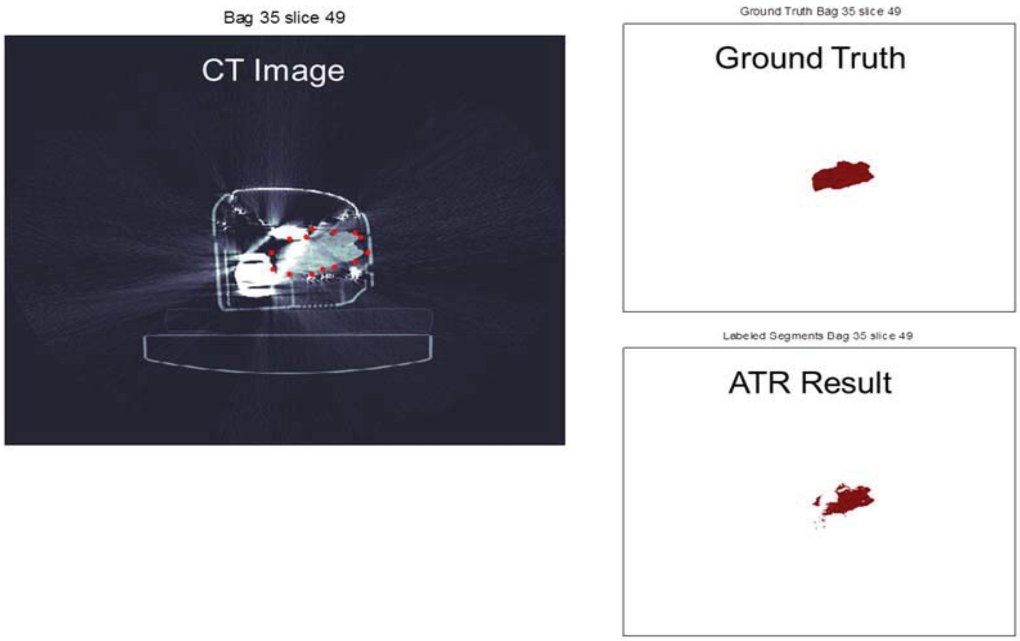
Case #2: Bulk with bad shading caused by beam hardening and scatter



Detected: YES
Precision: 72.2%
Recall: 94.2%

The target is merged with the overlapping portion of the sheet on top of the target reducing the precision

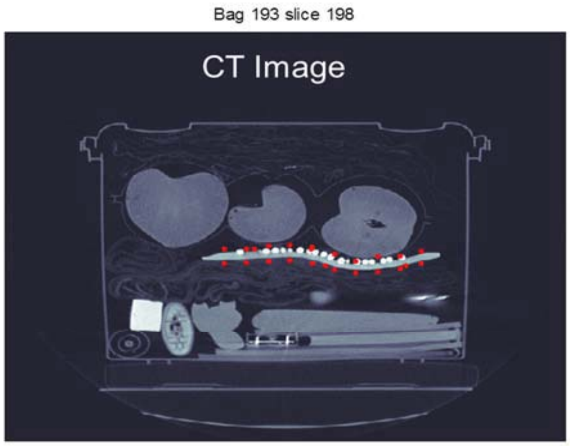
Case #3: Bulk inside electronics

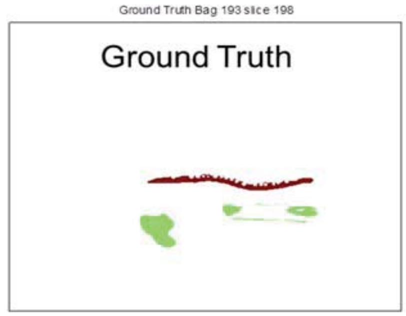


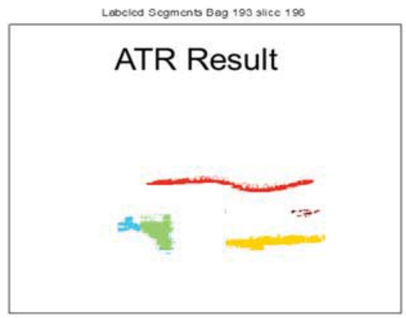
Detected: YES
Precision: 87.5%
Recall: 79.3%

Some parts of the target image are removed due to the metal artifacts

Case #4: Bulk with texture



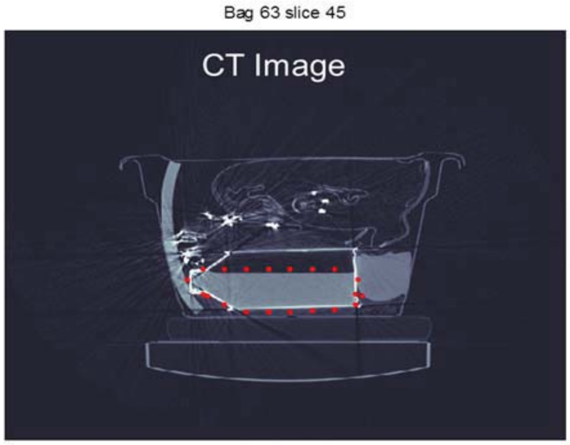


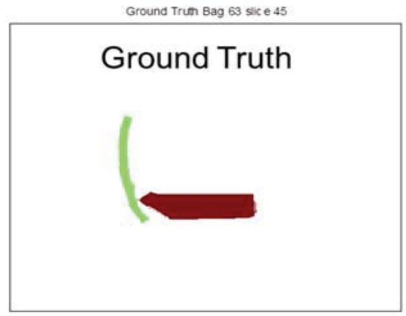


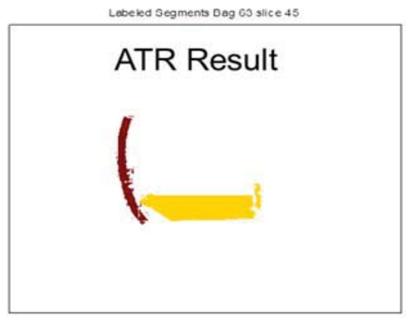
Detected: YES
Precision: 96.5%
Recall: 73.8%

The ground truth label identifies some of the texture as part of the target. The ATR removes most of them resulting in a lower recall

Case #5: Bulk with density close to water (~5% saline)



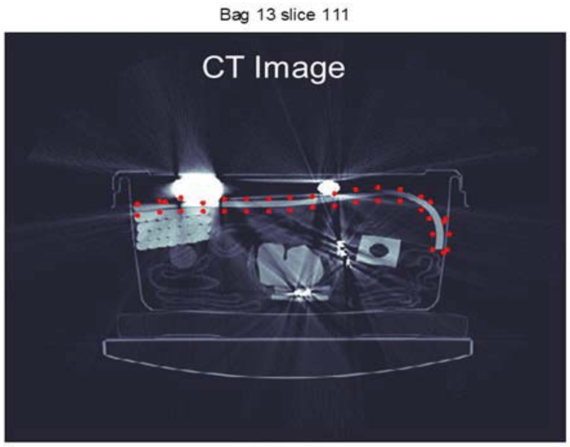


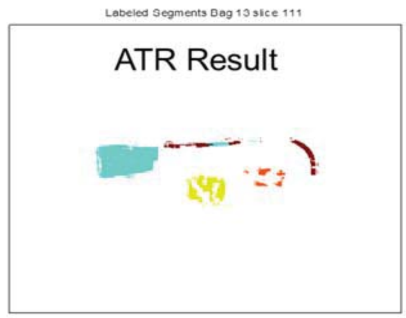


Detected: YES
Precision: 93%
Recall: 95.5%

Object was detected with no issues

Case #6: Sheet with bad streaks caused by metal, beam hardening and scatter

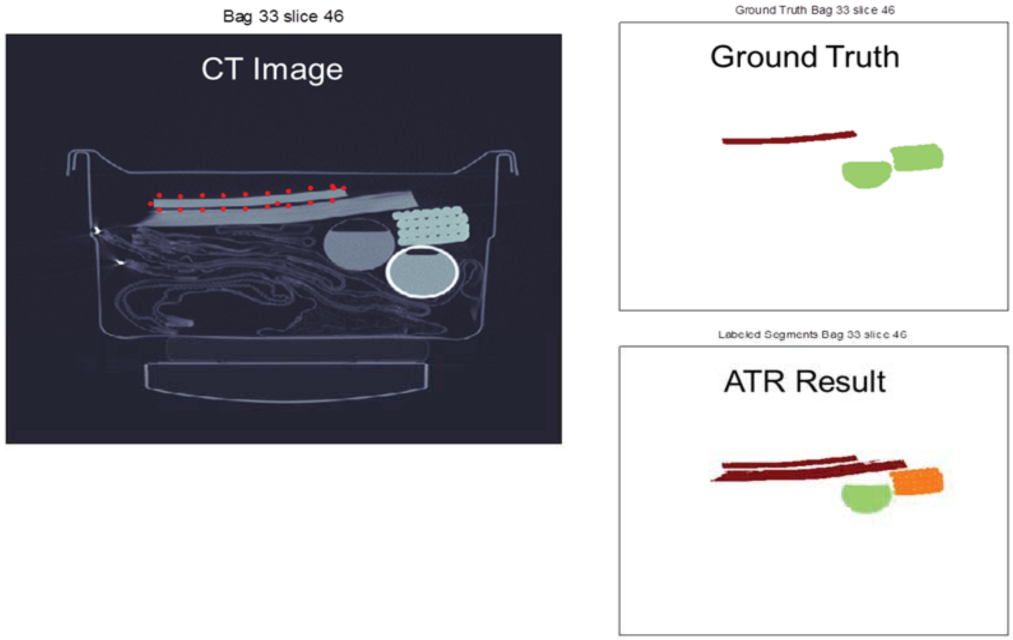




Detected: YES
Precision: 83.3%
Recall: 26.7%

The sheet was split into multiple pieces resulting in a low recall.

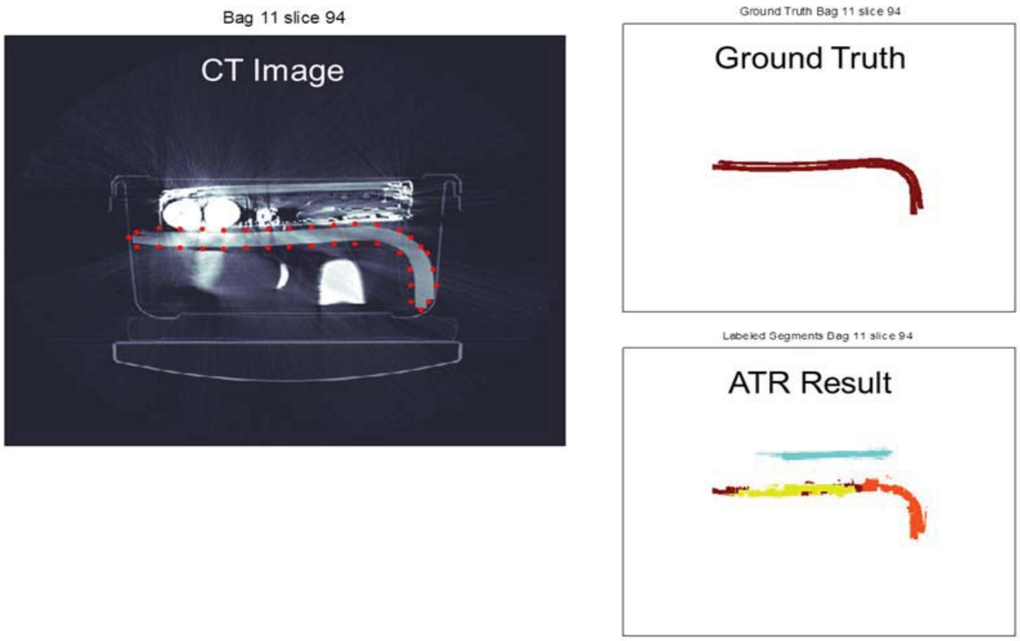
Case #7: Sheet laying on top of another flat object



Detected: YES
Precision: 21.1%
Recall: 82.7%

The sheet is detected fully but is merged with the object below it, though still met the test criteria. In this plane the objects are separated but in other planes they appear merged. A more sophisticated splitting algorithm based on image voxel data instead of voxel slabs probably would have separated the two objects, but such an algorithm was not necessary to meet the requirements of this test.

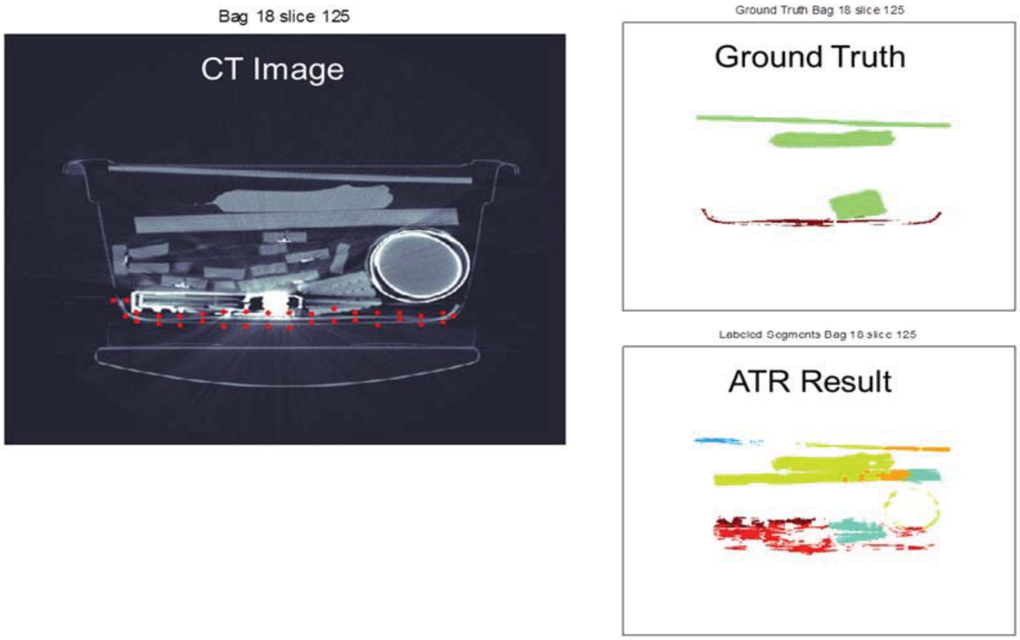
Case #8: Object with lots of photon starvation



Detected: YES
Precision: 71.9%
Recall: 44.4%

The sheet is split into multiple objects reducing the recall of the object. Improved merging algorithms would probably be capable of merging the object but were not necessary for this test. Note also there is a false alarm object in this frame.

Case #9: PT sheet based on thickness

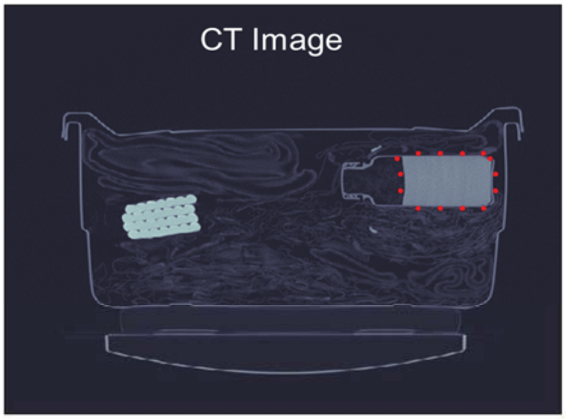


Detected: YES
Precision: 23.2%
Recall: 32.6%

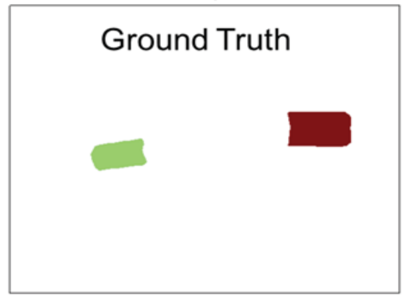
This sheet is not particularly well captured but sufficiently to count as a detection in this test. The sheet itself appears merged with some of its surrounding objects.

Case #10: PT Powder (based on density, not mass)

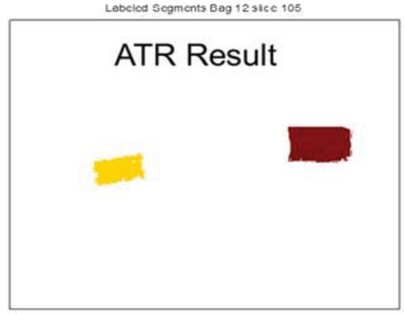
Bag 12 slice 105



Ground Truth Bag 12 slice 105



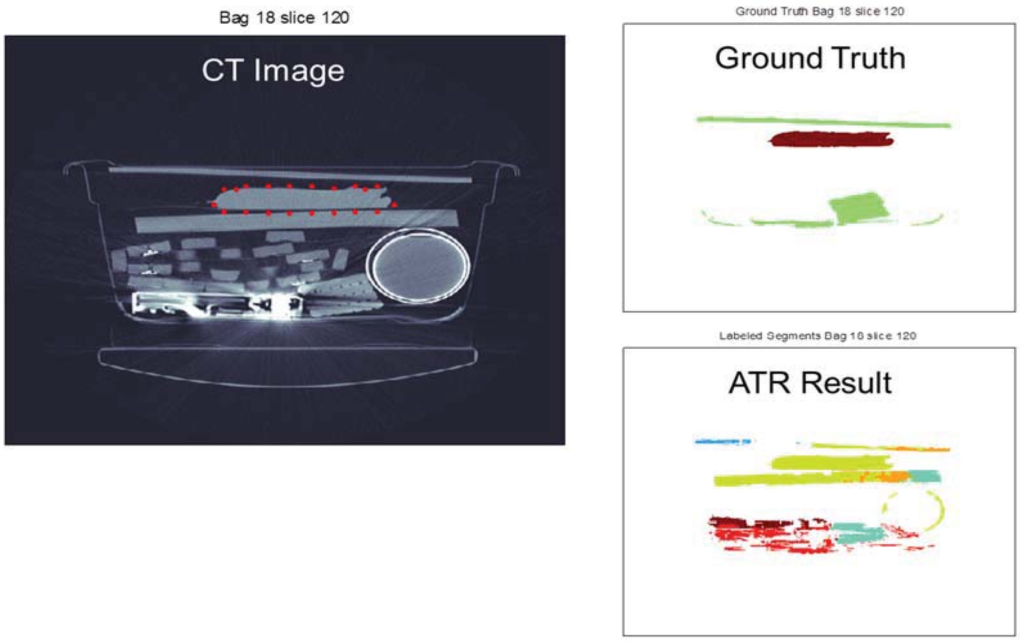
Labeled Segments Bag 12 slice 105



Detected: NO
Precision: 49.95%
Recall: 96%

Powders were not considered detection requirements in the final version of the test so the powder detector was mostly turned off, though detection did not count against the final score so some were detected. In this case the object is merged with another object of similar size and shape behind the frame in this image so it is not visible.

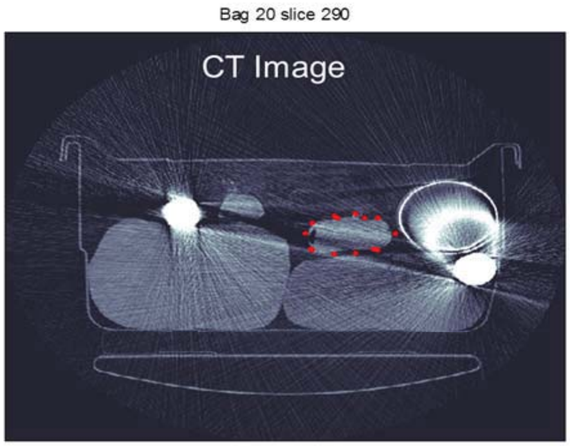
Missed Detection #1: Merger

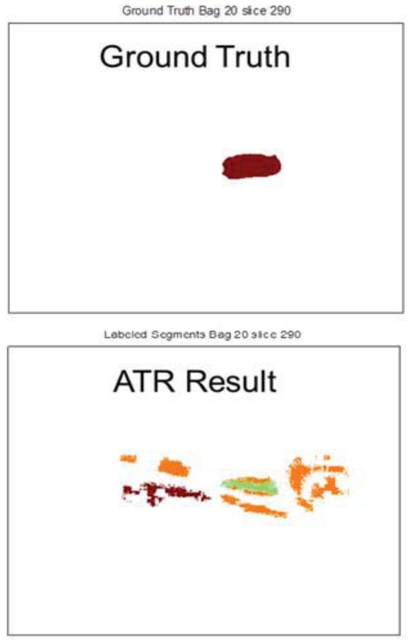


Detected: NO
Precision: 28.1%
Recall: 90.4%

The object in question is merged with a non-target object of similar density below it in the image.

Missed Detection #2: Metal artifacts





Detected: NO
Precision: 23.0%
Recall: 38.4%

The object in question is badly distorted by metal artifacts so the object itself was split into multiple pieces with different apparent densities, and the ATR did not piece enough of it together to count as a detection.

There were a total of 13 missed targets after the split and merge stage the exact specifications are shown in following table

Bag	Target	type	Prec	recall
13	6047	R	92	47
15*	6045	C	98	46
16*	6002	S	33	97
18*	6025	S	28	90
18*	6051	C	79	32
18*	8031	R sh	5	17
20	6012	S	23	38
34	6012	S	95	43
38	6001	S	41	98
115	6178	S	46	92
147*	6140	R sh	18	65
162*	6573	R sh	15	93
183	6557	S	20	65

- objects were detected in some tests but not final results

4 missed detections due to splitting
7 missed detections due to merging
2 missed detections had both split and merge issues