# Implementing Mahalanobis Distance Computation on GPUs for portability

Nelián E. Colón-Collazo, Undergraduate Student
Advisors: Nayda G. Santiago, Miguel Vélez-Reyes
{nelian.colon, naydag.santiago, miguel.velez7} @upr.edu

## Abstract

Hyperspectral imaging (HSI) has been identified as an important sensing technology for trace detection of explosives on surfaces such as clothing, packaging and skin. HSI contain more information than black and white or color images because they contain data across the electromagnetic spectrum at high spectral resolution. The biggest challenge with HSI is the large amount of data that needs to be processed in order to meet real-time requirements in applications such as portal or stand-off detection systems. To improve the performance and throughput of HSI-based detection systems, General-Purpose computation on GPUs has been proposed as a solution. GPUs' parallel architecture makes it well suited for data parallel processes which arise in many HSI processing applications. NVIDIA's CUDA provides an API that allows us to take full advantage of the parallelism in the GPU. The goal of the project is to implement algorithms for HSI processing in GPUs and provide an open source portable library that can facilitate development of HSI-based detection systems. Strategies for processing all the data and, at the same time ensure portability, are being explored. This work discuss memory management techniques and preliminary results for the computation of Mahalanobis Distance used in HSI classification algorithms.

## Relevance

This work addresses the use of GPUs high-throughput HSI processing which is required in portal and stand-off explosives detection systems. The proposed library will provide a set of tools for HSI-processing rapid prototyping.

## Mahalanobis Distance

• The Mahalanobis Distance (MD) is used to measure the separation of two groups of objects.

• It is a useful way of determining similarity of an unknown sample set to a known one. It can also be defined as dissimilarity measure between two random vectors of the same distribution with the covariance matrix.

• It differs from Euclidean distance in that it takes into account the correlations of the data set.

• The sample MD is defined as: $D_M^2(x) = (x-\mu)^T S^{-1}(x-\mu)$

- Where x is the data, μ is the mean, (x - μ) is the mean difference vector, $(x-\mu)^T$ is the transposed mean difference vector, and $S^{-1}$ is the inverse of the pooled covariance matrix.

•MD is widely used in cluster analysis and classification techniques.

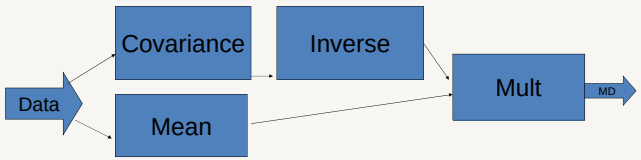


*Fig 1: Main components of the computation of Mahalanobis distance.*

## Technical Approach

• Approach
- Partition and send various kernels for each partition [3].
- Computation of sample covariance and mean are parallel
- Inverse is inherently serialized.
- Matrix multiplication parallelizes well.

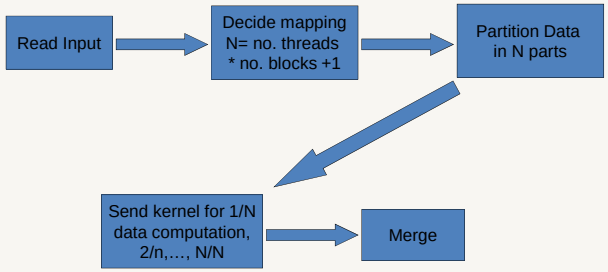


*Fig 2: Managing memory limitations*

- Two different approaches for processing the data for portability
  - Map the number of threads and blocks according to GPU architecture.
    - Pro: Takes advantage of GPU capabilities.
    - Con: User input required to know the type of GPU.
  - Thread and block mapping independent of GPU.
    - Pro: Simplifies code and requires no user input.
    - Con: Underuse of GPU resources.

## Accomplishments Through Current Year

- Learn about Mahalanobis Distance
- Learn about memory management
- Complete implementation of the Mahalaobis Distance in C.
- Implement MD in CUDA to use the GPUs; needs refinement for larger datasets.

## Future Work

Finish the implementation of the Mahalanobis Distance in CUDA, incorporating memory management and data processing techniques.

## Opportunities for Transition to Customer

This may be incorporated in products designed for HSI processing in defense and security applications.

## References

1. *Kirk, D., & Hwu, W. (2010). Programming Massively Parallel Processors. Burlington, MA: Elsevier.*
2. *Jenness, J. (2003). Mahalanobis Distances Description.* http://www.jennessent.com/arcview/mahalanobis_description.htm *(Fig. 1)*
3. *"Partitioning: Domain Decomposition". Designing and Building Parallel Programs* http://www.mcs.anl.gov/~itf/dbpp/text/node16.html

This material is based upon work primarily supported by the U.S. Department of Homeland Security under Award Number 2008-ST-061-ED0001. 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.