Activity Recognition using Dynamic Subspace Angles

Octavia I. Camps, Mario Sznaier; Students: Binlong Li, Mustafa Ayazoglu, Teresa Mao, and Necmiye Ozay Dept. of Electrical and Computer Engineering, Northeastern University camps@coe.neu.edu, msznaier@coe.neu.edu

Northeastern

Robust Systems Lab

Abstract

Cameras are ubiquitous and can significantly change the way we live and interact with our environment. **Human activity recognition is central to understanding dynamic scenes for applications ranging from surveillance, to assisted living, to video gaming without controllers.**

In this research, we propose a novel approach for activity recognition by comparing the principal angles of dynamic subspaces, which are found by performing SVDs of the experimental data.

The proposed approach outperforms state-of-the-art methods in very complex scenarios involving interacting actors, as well as in the classical KTH benchmark dataset.

Relevance

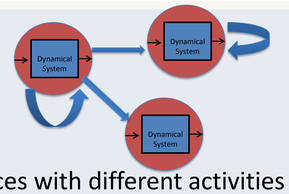
Automatic activity recognition is key to detect threats and suspicious behaviors in video sequences.

Current state-of-the-art methods for activity recognition are based on space-time interest points extracted at the frame level or visual code words. The lack of use of dynamics information limits their ability to recognize long and complex actions.

In contrast, our new approach to automatic behavior understanding exploits the temporal information encoded in the data by modeling activities as the output of unknown dynamic systems evolving from unknown initial conditions. As a result, the proposed method is capable of parsing and recognizing complex activities and human interactions in long video sequences.

Technical Approach

Main Ideas:



•Long sequences with different activities are modeled as hybrid systems with switches, mathematically described by graphs. Single activities are modeled as linear dynamical systems.

•Switches between activities can be found by semidefinite programming.

•Output trajectories can be compared in terms of dynamic subspaces angles (DSA), defined as the canonical correlations between the subspaces spanned by their Hankel matrices.

•Discriminant canonical correlations (DCC) can be used to maximize the canonical correlations of with-in classes while minimizing the canonical correlations of between-classes

Parsing Activities:
Enables Detection of Suspicious Behaviors

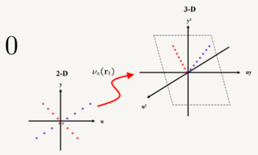
•Given a sequence, the number of systems, and noise bounds, find a piecewise affine system such that: $0 = \mathbf{b}(\sigma_t)^T \mathbf{r}_t + \eta_t$

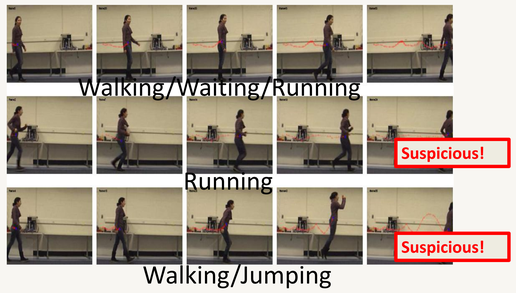
• Can be solved using polynomial optimization on the Veronese map of the data.

$$p_s(\mathbf{r}) = \Pi_{i=1}^{s}(\mathbf{b}_i^T \mathbf{r}_t) = 0$$

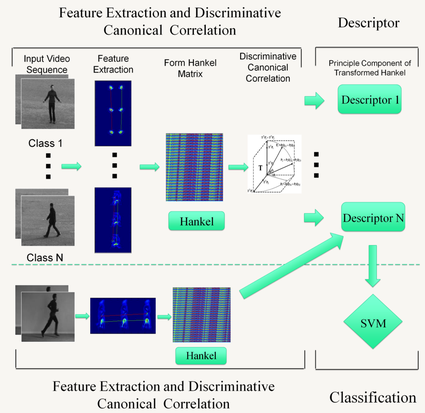
$$\min_{m^{(t)}} \operatorname{rank} \tilde{\mathbf{V}}_s(\mathbf{r}_t, m^{(t)})$$

subject to: each $m^{(t)}$ is a moment sequence





Activity Recognition:



•Activities are represented by the subspaces of their feature (HOGs, cuboids, etc.) trajectories.

•The subspaces are found by performing a SVD on the Hankel matrix of the measurements.

•DCC is used to align subspaces of sequences from the same activity to increase robustness to variations between actors.

•A support vector machine (SVM) is used to classify activities according to their DSAs.

Feature Extraction--HoG

- HoG, each actor
- 16 x 8 grid

Algorithm	Accuracy on TV dataset
Ours (2010)	68%
Patron-Perez et al (2009)	54.45%

Algorithm	Accuracy on KTH dataset
Ours (2010)	93.6%
Wang et al (2009)	92.1%
Laptev et al (2008)	91.8%

Accomplishments Through Current Year

•Robust activity parsing allows to detect contextually abnormal sequences of activity.

•24% accuracy improvement over the state-of-the-art human interaction recognition

Future Work

•Use tracklets to handle occlusions and tracking failures.

•Incorporate dynamical time warping.

•Extend to more than 2 actors and more complex activities.

Opportunities for Transition to Customer

Understanding human activities and interactions from video has application in visual surveillance, assisted living, human computer interfaces and entertainment.

The proposed algorithms have been prototyped in MATLAB and are simple to implement using C++.

Patent Submissions

None yet.

Publications Acknowledging DHS Support

N. Ozay, M. Sznaier, O. Camps and C. Lagoa, *"GPCA with Denoising: A Moments-Based Convex Approach,"* CVPR, June 2010.
N. Ozay, M. Sznaier and C. Lagoa, *"Model (In)validation of Switched ARX Systems with Unknown Switches and its Application to Activity Monitoring,"* 2010 IEEE CDC.
N. Ozay, M. Sznaier, C. Lagoa and O. Camps, *"A Sparsification Approach to Set Membership Identification of a Class of Affine Hybrid Systems,"* IEEE Trans. Aut. Contr., conditionally accepted for publication, 2011.
B. Li, M. Ayazoglu, T. Mao, O. Camps and M. Sznaier, "Activity Recognition using Dynamic Subspace Angles," 2011 IEEE Conf. on Computer Vision and Pattern Recognition (CVPR). To appear.

Other References

Patron-Perez, M. Marszalek, A. Zisserman, and I. Reid. High five: Recognising human interactions in tv shows. In BMV Conference, 2010.
H. Wang, M. Ullah, A. Klaser, I. Laptev, and C. Schmid. Evaluation of local spatiotemporal features for action recognition. In BMV Conference, pages 1–11, 2009.
Laptev, M. Marszalek, C. Schmid, and B. Rozenfeld. Learning realistic human actions from movies. CVPR 2008

This material is based upon work 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.