Human Action Recognition using Hankel Matrices

Teresa Mao, Octavia Camps, Mario Sznaier
Electrical and Computer Engineering, Northeastern University, Boston, MA

Robust Systems Lab

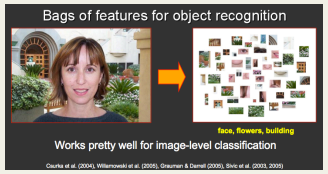
Abstract

Surveillance systems are used everywhere, but the manpower required to monitor and analyze surveillance videos is often expensive. Human activity recognition is another step towards developing better surveillance systems.

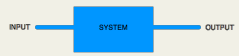
Applications
- surveillance in public spaces
- pedestrian on the streets
- fall detection for older people
- film and media analysis

Overview

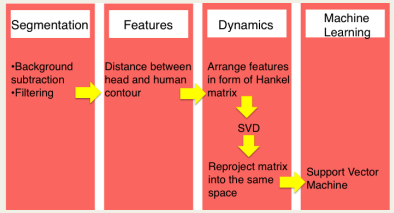
- Most of the current human activity recognition algorithms are carried out at the frame level and neglect the temporal data association between the frames.
- For example, bags of features cannot detect events distinguished by temporal information (i.e. standing up and sitting down)



In our approach, we view video sequences as the output of dynamic systems and use the Hankel matrix and its properties to help us exploit the data association between video frames.

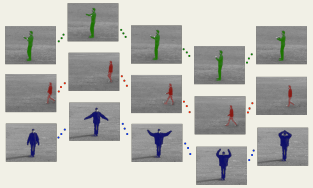


Algorithm



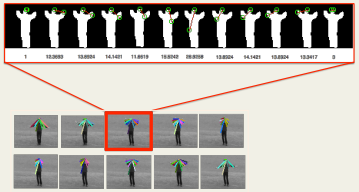
Segmentation: Human Detection

- Average filter
- Background subtraction



Human Shape Features

- Length from the reference point (head) to the contour at every 30 degree angles
- For each image, we describe the human shape with 12 length measurements. For each video, there are 30 frames, length matrix is 30 x 12
- In the actual implementation, we rotate each image 30 degrees and measure the length from reference point to the contour at 0 degrees



Dynamics

Hankel

$$\mathrm{H}_f(k,l) \doteq \begin{bmatrix} \mathrm{f}_1 & \mathrm{f}_2 & \cdots & \mathrm{f}_l \\ \mathrm{f}_2 & \mathrm{f}_3 & \cdots & \mathrm{f}_{l+1} \\ \vdots & \vdots & \ddots & \vdots \\ \mathrm{f}_k & \mathrm{f}_{k+1} & \cdots & \mathrm{f}_{k+l-1} \end{bmatrix}$$

- Definition: Hankel matrix is a matrix with constant positive sloping skew-diagonals.
- The rank of Hankel matrix represents the dynamic order of the video sequence.

Dynamics (cont.)

- The Hankel matrix allows you to arrange the length vector for each frame into partially overlapping segments and rearranging them into a matrix

Single Value Decomposition

- Concatenate all Hankel matrices into one matrix.

$$H_{all} = \begin{bmatrix} H_{boxing} \\ H_{waving} \\ H_{clapping} \\ \vdots \end{bmatrix}$$

- The SVD of the Hankel matrix can still provide the dynamic order of the Hankel matrix.

$$H_{all} = U \Sigma V$$

Reprojection

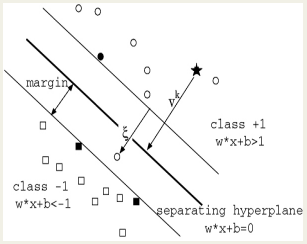
- The U Matrix is especially useful to us to "align" matrices.
- To make activities more comparable, matrices are reprojected onto the same subspace using the U matrix.

$$H_{reprojected} = U^T * H_{all}$$

- Find the energies on each direction by looking at the norms of the rows of $H_{reprojected}$

Supervised Learning: SVM

- Human activity recognition using SVM is a new method in this area of research.
- The purpose of SVM is to find an optimal hyperplane under the linear separable condition. As for nonlinear separable conditions, it uses a kernel to map between the features spaces.
- This optimal hyperplane leaves the largest possible fraction of points of same class on the same side



Results

Previous work:

	HOG3D	HOG/HOF	HOG	HOF	Cuboids	ESURF
Harris3D	89.0%	**91.8%**	80.9%	**92.1%**	–	–
Cuboids	**90.0%**	88.7%	82.3%	88.2%	89.1%	–
Hessian	84.6%	88.7%	77.7%	88.6%	–	81.4%
Dense	85.3%	86.1%	79.0%	88.0%	–	–

Our method: tested clapping, waving and boxing. Using only decent segmentations.

Average accuracy rate = 96.3%

Future work

- Expand into multiple action detection or event detection
- Less segmentation dependent
- Implementation for real time applications

References

Schölkopf ,B., Burges, C., Smola, A., *Geometry and Invariance in Kernel Based Methods*, in Advances in Kernel Methods - Support Vector Learning, MIT Press, Cambridge, USA, 1999

Tao Ding; Sznaier, M.; Camps, O.I.; *Fast Track Matching and Event Detection*. CVPR 2008

W.Ding, H. Li, X. Li, J. Zhe, *Unsupervised Spatio-temporal Multi-Human Detection and Recognition in Complex Scene*. Image and Signal Processing. 2009

Yu, A. C. H.; Cobbold, R. S. C.; *A New Eigen-Based Clutter Filter Using the Hankel-SVD Approach*, Ultrasonics Symposium, 2006. IEEE

Conclusion

- View video sequences as dynamic systems and use the Hankel matrix and its properties to help us understand the data association between video frames.
- We train and test with a popular supervised learning method (Support Vector Machine) for human activity recognition.
- Explored basic human activities such as clapping, waving, boxing and walking from the KTH dataset, our results are very promising.

"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."