Canonical Correlation Analysis of Dynamics for Human Action Categorization

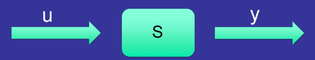
Binlong Li, Octavia Camps, Mario Sznaier
Electrical and Computer Engineering, Northeastern University, Boston, MA 02115

INTRODUCTION

Current state-of-art methods for activity recognition are based on space-time interest points extracted at the frame level or visual code words without any dynamics information. Our proposed method is based on canonical correlation analysis and the Hankel matrices which encapsulates the underlying dynamics of video sequences.

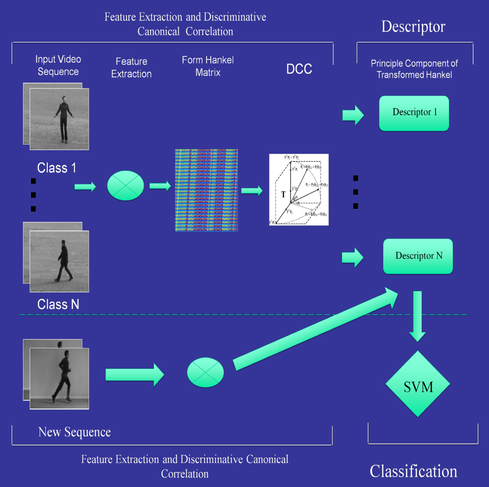
XBOX360 will release Kinect, By using Motion recognition techniques, to realize Controller-free Game User Interface.

Proposed Approach



- Model activities using spatial-temporal features that are the outputs of linear dynamic systems
- Features from similar activities are generated by similar dynamic systems.
- Use DCC to better separate the activities.

PROCEDURE FLOW

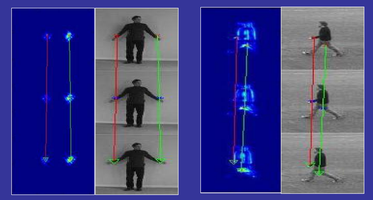


Feature Detection & Tracking

- 3D Gabor filter as Spatial-temporal detector

$$h_{ev}(t;\tau,\omega) = -cos(2\pi t\omega)e^{-t^2/\tau^2}$$
$$h_{od}(t;\tau,\omega) = -sin(2\pi t\omega)e^{-t^2/\tau^2}$$

- Tracking 2 regions with largest response through out 3D spatial-temporal detection.



(a) Hand Clapping (b) Walking

Figure 3. Tracking illustration. For Each figure, left is response images, right is original images.

Discriminative Canonical Correlations

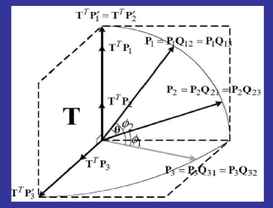


Figure 5. Conceptual illustration of DCC. P1 and P2 are within-class sets and P3 is other classes. After projection by T, learned from DCC, the within-class distance has been minimized and between-class sets are maximized.

$$\mathbf{T} = \max_{argT} \mathrm{tr}(T^T S_b T)/\mathrm{tr}(T^T S_w T)$$

$$S_b = \sum_{i=1}^{m} \sum_{l \in B_i} (P_l' Q_{li} - P_i' Q_{il})(P_l' Q_{li} - P_i' Q_{il})^T$$

$$S_w = \sum_{i=1}^{m} \sum_{k \in W_i} (P_k' Q_{ki} - P_i' Q_{ik})(P_k' Q_{ki} - P_i' Q_{ik})^T$$

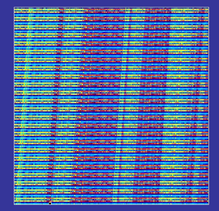
$$B_i = \{j \mid X_j \notin C_i\} \qquad W_i = \{j \mid X_j \in C_i\}$$

$$H_i \rightarrow H_i' = T^T H_i$$

Hankel Construction

- Hankel is a matrix with constant skew diagonals
- Rank of Hankel Matrix can capture the order of a linear dynamic system
- each v_i represents the frame feature extracted from frame i

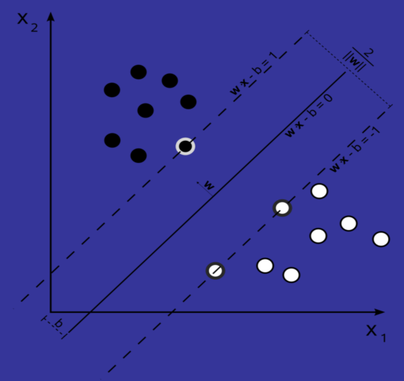
$$H = \begin{bmatrix} v(1) & v(2) & \cdots & v(n-m+1) \\ v(2) & v(3) & \cdots & v(n-m+2) \\ \vdots & \vdots & \ddots & \vdots \\ v(m) & v(m+1) & \cdots & v(n) \end{bmatrix}$$



(a) Hankel Matrix (b) Real Data

SVM

- Support vector machine constructs a hyperplane or set of hyperplanes, with largest margin to the nearest training data point, in a high dimensional space, which can be used for classification.
- New samples are then mapped into that same space and predicted to belong to a category based on which side of the gap they fall on.
- In recent years, SVM has been widely used by solving Computer vision related classification problems



Experiment Result

We tested our proposed approach on a public action data set (KTH), such as, boxing, clapping, waving, walking and running actions, comparing to the other state-of-art algorithms, our method showed a promising performance.

	Box	Clap	Wave	Walk	Run
Box	0.92	0.00	0.00	0.06	0.03
Clap	0.14	0.86	0.00	0.00	0.00
Wave	0.06	0.06	0.89	0.00	0.00
Walk	0.00	0.00	0.00	0.97	0.03
Run	0.00	0.00	0.03	0.03	0.94

Method	Accuracy Rate(%)
Our Method	91.67
Dollar et al.	81.17
Schuldt el al.	71.72
Ke et al.	62.96
Jun et al.	82
Niebles et al.	91.3

Conclusions

Our experiment shows that temporal information between spatial-temporal interesting points have critical impact on the action recognition problems.

Literature cited

Juan Carlos Niebles, Chih-Wei Chen, and Li Fei-Fei. Modeling temporal structure of decomposable motion segments for activity classication. ECCV, pages 1-14, Jul2010.

Tae-Kyun Kim, J Kittler, and R Cipolla. Discriminative learning and recognition of image set classes using canonical correlations. Pattern Analysis and Machine Intelligence, IEEE Transactions on DOI - 10.1109/TPAMI.2007.1037, 29(6):1005-1018, 2007.

Acknowledgments

"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."