Histogram of Oriented Displacements (HOD): Describing Trajectories of Human Joints for Action Recognition

Mohammad A. Gowayyed[1], Marwan Torki[1], Mohamed E. Hussein[1], Motaz El-Saban[2]

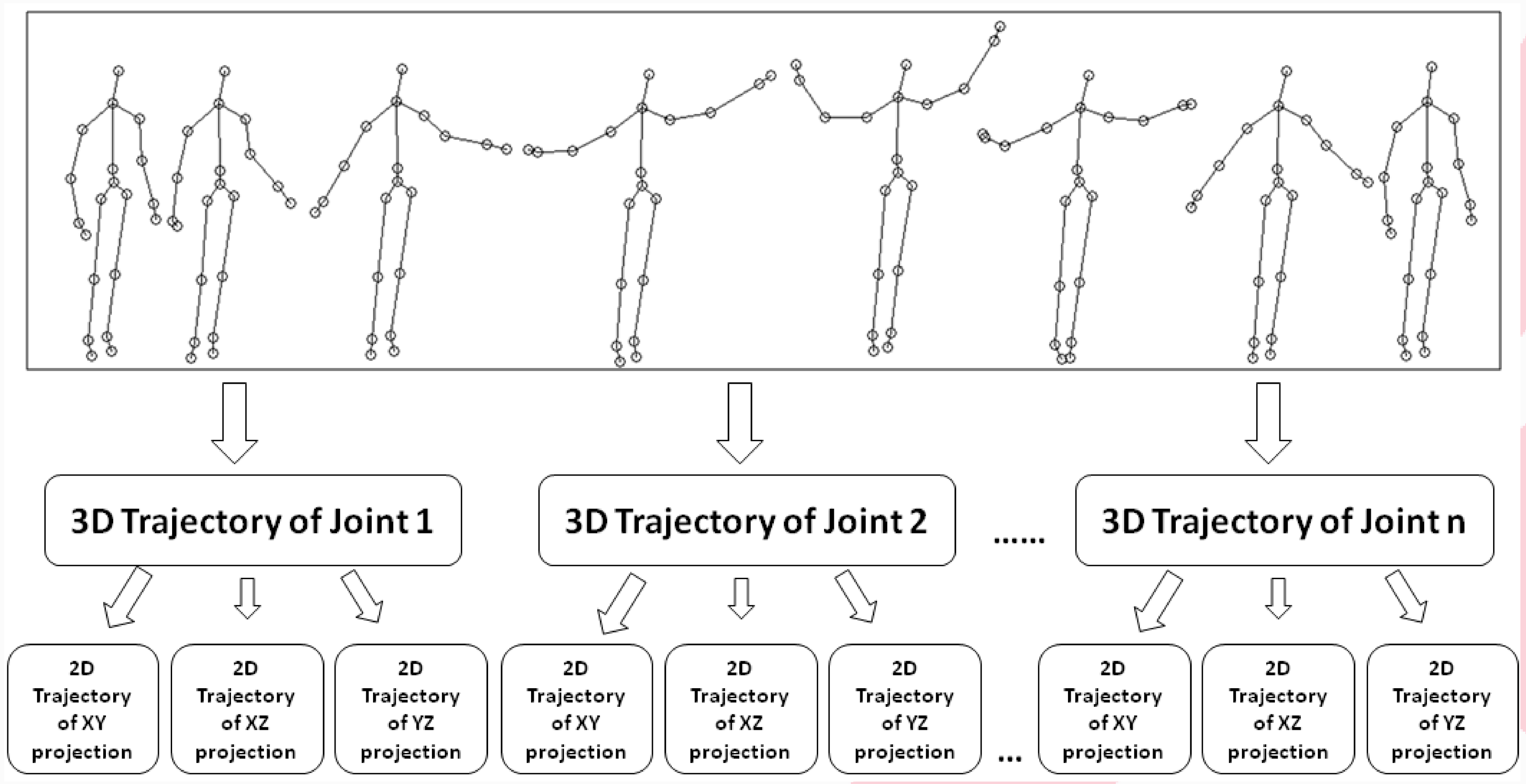
1 Alexandria University

2 Microsoft Research, Advanced Technology Labs Cairo

IJCAI-13

Abstract - Creating descriptors for trajectories has many applications in robotics/human motion analysis and video copy detection. Here, we propose a novel descriptor for 2D trajectories: Histogram of Oriented Displacements (HOD). Each displacement in the trajectory votes with its length in a histogram of orientation angles. 3D trajectories are described by the HOD of their three projections. We use HOD to describe the 3D trajectories of body joints to recognize human actions, which is a challenging machine vision task, with applications in human-robot/machine interaction, interactive entertainment, multimedia information retrieval, and surveillance. The descriptor is fixed-length, scale-invariant and speed-invariant. Experiments on MSR-Action3D and HDM05 datasets show that the descriptor outperforms the state-of-the-art when using off-the-shelf classification tools

Approach



Given a sequence of body joints locations of a human performing an action in n frames, our goal is to provide a discriminative descriptor for this sequence. We describe the 3D trajectory of each individual joint, then concatenate the descriptors of all joints to form the final descriptor. Each 3D trajectory is represented by the HOD of its three 2D projections (xy, xz and yz).

Histogram of Oriented Displacements (HOD)

For a 2D trajectory where P_t is the position of the joint at time t. The trajectory is described by a histogram of n bins (say 8 in Fig. 2).

For each displacement, the angle θ and the length of the displacement are calculated. The length is added to the appropriate histogram bin.

$$histogram\_bin = \frac{angle * histogram\_length}{360}$$

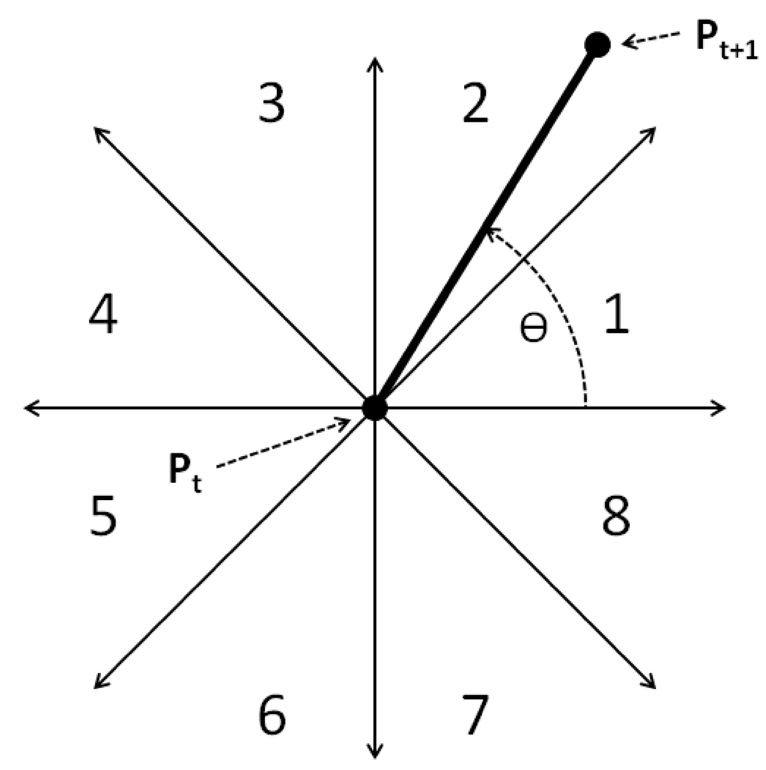
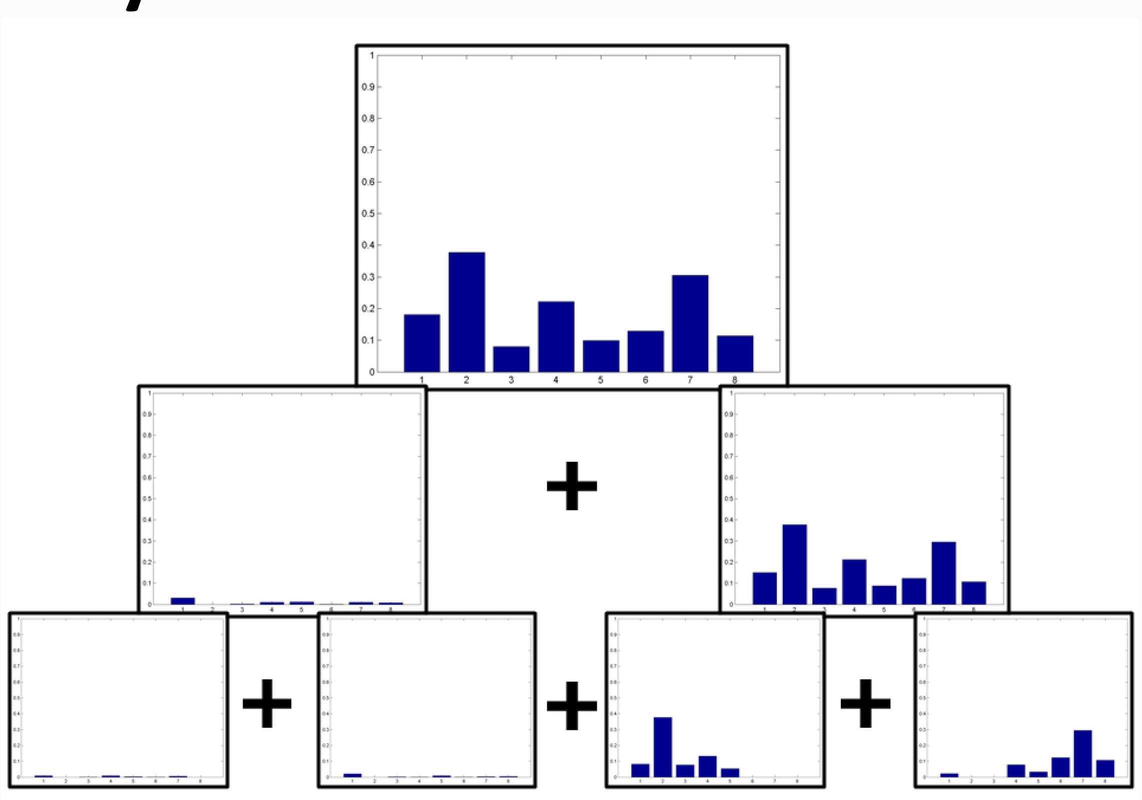


Figure 2: For the displacement shown in this figure, the length of (P_t, P_{t+1}) will be added to the second histogram bin.

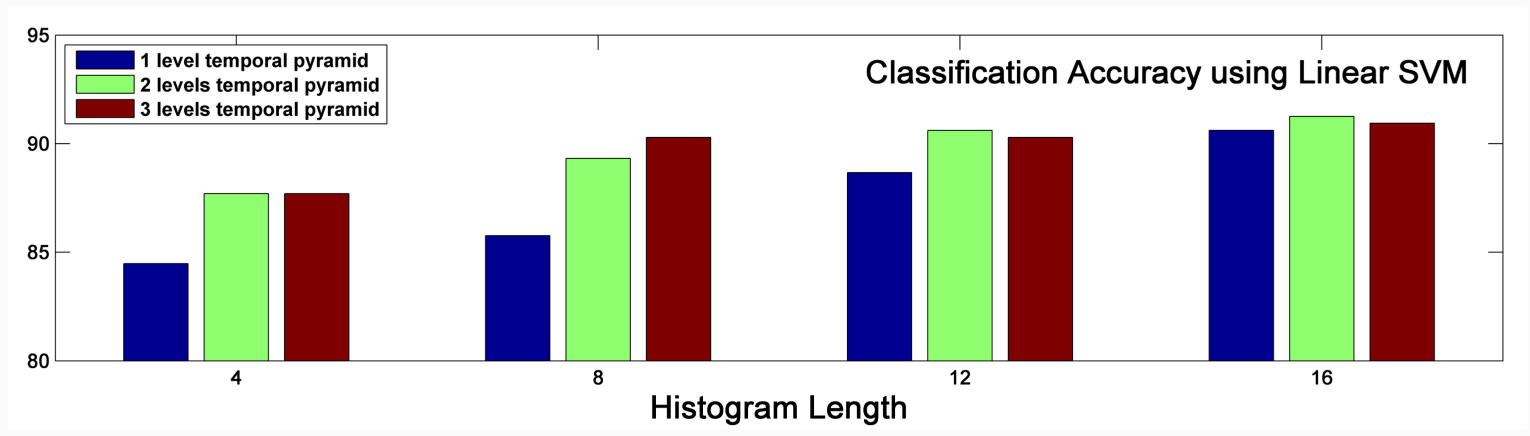
Temporal Pyramid



Results

MSR-Action3D

Method	Accuracy (%)
Recurrent Neural Networks [1]	42.5
Hidden Markov Model [2]	78.97
Action Graph on Bag of 3D Points [3]	74.7
Random Occupancy Patterns [4]	86.5
Actionlets Ensemble [5]	88.2
2-level 16-bin HOD (20 joints)	**91.26**
2-level 16-bin HOD (right hand joint only)	**74.07**
1-level 4-bin HOD (weakest configuration)	**84.47**



HDM05

Method	Accuracy (%)
Sequence of Most Informative Joints [6]	84.4
3-level 4-bin HOD (20 joints)	**97.27**
3-level 8-bin HOD (right elbow joint only)	**82.72**

References

[1] J. Martens and I. Sutskever. Learning Recurrent Neural Networks with Hessian-Free Optimization. In ICML, 2011.
[2] F. Lv and R. Nevatia. Recognition and Segmentation of 3-D Human Action Using HMM and Multi-class AdaBoost. In ECCV, 2006.
[3] W. Li, Z. Zhang, and Z. Liu. Action recognition based on a bag of 3d points. In Human Communicative Behavior Analysis Workshop (in conjunction with CVPR), 2010.
[4] J. Wang, Z. Liu, J. Chorowski, Z. Chen, and Y. Wu. Robust 3d action recognition with random occupancy patterns. In ECCV, 2012.
[5] J. Wang, Z. Liu, Y. Wu, and J. Yuan. Mining actionlet ensemble for action recognition with depth cameras. In CVPR, 2012.
[6] F. Ofli, R. Chaudhry, G. Kurillo, R. Vidal, and R. Bajcsy. Sequence of the most informative joints (SMIJ): A new representation for human skeletal action recognition. In International Workshop on Human Activity Recognition from 3D Data, 2012.