Histogram of Body Poses and Spectral Regression Discriminant Analysis for Human Action Categorization

Ling Shao
ling.shao@sheffield.ac.uk

Xiuli Chen
elp09xc@sheffield.ac.uk

Department of Electronic and Electrical Engineering,
The University of Sheffield

This paper explores a recently proposed and rarely reported subspace learning method, Spectral Regression Discriminant Analysis (SRDA) [1, 2], on silhouette based human action recognition. The recognition algorithm adopts the Bag of Words (BoW) model combined with the action representation based on Histogram of Body Poses sampled from silhouettes in the video sequence. In addition, we compare the performance of SRDA for dimensionality reduction with several traditional subspace learning methods, such as Principle Component Analysis (PCA), supervised Locality Preserving Projections (LPP), unsupervised LPP and Neighbourhood Preserving Embedding (NPE). Experimental results show that Histogram of Human Poses combined with SRDA or its kernel version, SRKDA, can achieve 100% recognition accuracy for the Weizmann human action dataset, which is better than any published results on the same dataset.

In this paper, we aim to investigate SRDA and SRKDA for dimensionality reduction in silhouette based human action recognition and compare its performance with other subspace learning methods. The main contribution of our work is listed below:
(1) We apply Spectral Regression for dimensionality reduction, including SRDA and SRKDA, on silhouette based human action recognition and compare the performance with that of other subspace learning methods, such as PCA, LPP, and NPE.
(2) Histogram of Body Poses (HBP) is proposed to describe a human action sequence. Only raw data sampled from the silhouettes associated with the video sequences are used to represent human poses.

Based on our experimental results, SRDA and SRKDA are proved to outperform other subspace algorithms in terms of dimensionality reduction for human action recognition. The integration with the HBP descriptor achieves better recognition accuracy than any other existing method.

The methodology used for human action recognition in this work is an extension of the Bag of Words (BoW) model to the visual domain, which is called Bag of Features or Bag of Visual Words. Similar to the BoW model in text classification, the first step is to build a visual vocabulary (also called codebook). One simple way is clustering all the feature vectors obtained from all the training samples by k-means clustering. The centre of each cluster is then defined as a codeword, and the size of the visual dictionary is therefore the number of the clusters. Given N vectors $[x_1; x_2; \cdots x_N]$, we perform k-means clustering to form k clusters, then we can find k cluster centers , i.e. codewords $[w_1, w_2 \cdots w_k]$. Each vector in $[x_1; x_2; \cdots x_N]$ can be assigned to its closest codeword. Each action represented by certain number of vectors can be described as the probability distribution of the codewords in the form of histogram.

In interest point based action recognition, each feature vector is a descriptor calculated around a detected interest point in an action sequence. In this paper, the features we use are body poses instead. Body poses, represented as normalized silhouettes for which the dimensionality is reduced using a certain subspace learning algorithm, from each frame in the training action sequences are fed into the BoW model. After k-means clustering, each action sequence can be described as a histogram of the occurrence of certain poses, which we call 'Histogram of Body Poses' (HBP). Examples of HBP are shown in Figure 1. The x-axis of the histogram is the codewords $[w_1, w_2 \cdots w_k]$ (k=40 here), and the y-axis value of w_i is the number of vectors that belong to w_i over the whole action sequence. These histograms are the data used to train the classifier, e.g. Support Vector Machine (SVM) or K Nearest Neighbours (KNN). During the testing stage, each test sequence is also represented as an HBP and fed into the classifier for action classification. The HBP descriptor is not only invariant to the length and periodicity of the action sequence but also robust to the swiftness and the starting posture of the action.

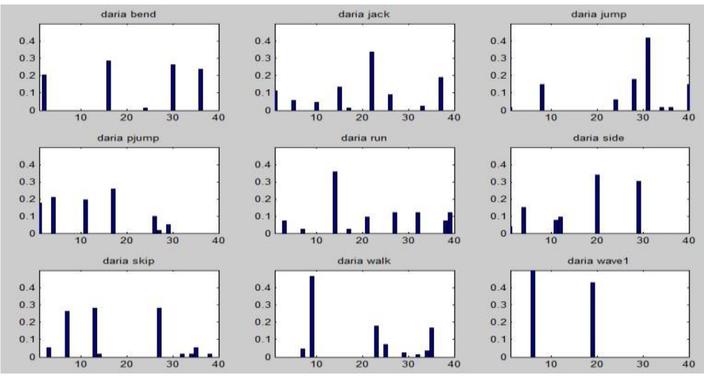

Figure 1: Several actions represented in the form of Histogram of Body Poses (k=40).

A widely used dataset, the Weizmann action dataset [3], is employed in our experiments. In practice, the scale and the position of the subjects often vary in different sequences due to changes in camera distance and viewpoint. Therefore, some pre-processing, including localization of the human body and normalization, should be done to make the silhouette cantered and the aspect ratio equal. Figure 2 shows the steps of the pre-processing.

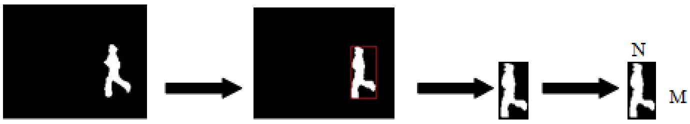

Figure 2: Pre-processing steps for the human silhouette.

In this paper, a novel silhouette-based human action recognition algorithm using Histogram of Body Poses for action representation and Spectral Regression Discriminant Analysis for dimensionality reduction is proposed. The HBP descriptor captures the probabilistic distribution of body postures and is invariant to the length and frequency of the performed action. SR-LDA and SR-KDA are proved to be superior to other subspace learning methods, such as PCA, LPP and NPE, for preserving the intrinsic structure of action silhouette data. The combination of HBP and SRDA achieves 100% accuracy on the popular Weizmann action dataset, which is the best performance ever reported.

[1] D. Cai, X. He, and J. Han. Regularized locality preserving indexing via spectral regression. In ACM International Conference on Information and Knowledge Management, November 2007.

[2] D. Cai, X. He, and J. Han. Spectral regression for efficient regularized subspace learning. In IEEE International Conference on Computer Vision (ICCV), October 2007.

[3] M. Blank, L. Gorelick, E. Shechtman, M. Irani, R. Basri, Actions as Space-Time Shapes. IEEE International Conference on Computer Vision, 17-21 Oct. 2005, vol. 2, 1395 –1402.