Action Recognition by Converting Normal Video into Silhouette Video

Divya Rajan N T[1] Blessingh T S[2] Dr. Sreeja Mole S S[3]
[1]ME Student [2]Assistant Professor [3]Head of the Department
[1]Department of Applied Electronics [3]Department of Electronics & Communication Engineering
[1,2,3]Narayanaguru College of Engineering, Manjalumoodu, Tamil Nadu

Abstract— Nowadays action recognition has many applications such as robotics, human computer interface etc. This project proposes a method for action recognition that uses a number of correlated poses. Here normal video is converting to silhouette video and then each silhouette video is converting to a set of images frames for feature extraction. The silhouettes are used as input features for the BoCP (Bag of Correlated Poses) model. BoCP are using for encoding the local features of action. After the descriptor extraction C-means clustering algorithm is using to generating the codebook. Clustering is the process of dividing the dataset into subsets, so that the data in each subset shares some common values according to some defined distance measure. To reduce the high dimensionality of computed feature Principal Component Analysis (PCA) and Linear Discriminant Analysis (LDA) are using. Experimental results prove the viability of the complementary properties of two descriptors and the proposed approach outperforms the state of the art methods on the IXMAS action recognition dataset.

Keywords: Bag of Correlated Poses, C-Means Clustering, Linear Discriminant Analysis, Principal Component Analysis

I. INTRODUCTION

Human action recognition has increasing attention over past years. But it is still a challenging problem due to difficulties such as background motion, occlusion and camera movement. Motion and structure are the main difficulties of the action occurring in a video sequence. Most of human motion analysis method is based on the assumption of view dependence.

Most of the current methods of action recognition are designed for limited view variations. View variations originate from the changing and frequently unknown positions of the camera. It is evident that such requirements on view dependence are difficult to achieve in realistic scenarios. One main difficulty lies in the fact that the motion field in an action region is separated by the background motion. The goal of this paper is to recognize human action in crowded environment by the use of the action models that are trained on the data with clean background. To satisfy this condition we need to develop techniques to obtain stable features that are able to recognize action but insensitive to background motion. Motion information includes the body or body part movements and position variation. To characterize an action we need to effectively encode both of them to obtain an informative representation.

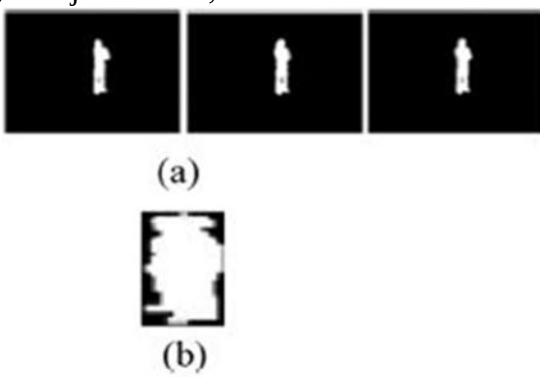

Fig. 1: Example of cross arm action

II. LITERATURE REVIEW

Due to the limitation of view dependent characteristic a large number of human motion analysis method had been kept away from adapting to a wider application spectrum. It is evident that the view point issue has been one of the bottlenecks for research development and practical implementation of human motion analysis. A large number of attempts and research progress on removal of the effect on human motion analysis method had been reported in recent years.

A. Multi View Human Action Recognition System Employing 2DPCA:

A novel for view invariant human action recognition is presented. This approach is based on Two Dimensional principal Component Analysis (2DPCA) and has many advantages when compared with 3 dimensional PCA. 2D-DCT can be used to compress the generated pattern and this algorithm is applied in both training and testing for feature extraction. This method reduces the computational complexity of the algorithm by maintaining the minimum storage requirements compared with other recent algorithm.

B. Selecting Key Poses On Manifold for Pair Wise Action Recognition:

In action recognition bag of visual words based approach is shown to be successful. This paper proposes a method for action recognition which models the descriptors by utilizing manifold learning technique to recover the geometric structure of the descriptor on a lower dimensional manifold. A page rank based centrality measure is using to select the key poses according to the recovered geometric structure. In each step a key poses selected from the manifold. Manifold learning is employed to generate the codebook. It not only reduces the dimensionality. Each action can be represented with a histogram of key poses.

III. PREVIOUS WORK

In ref (4) there proposed a method of action recognition using histogram of oriented gradient. It is based on the 3D extension of the HOG (Histogram of Oriented Gradient) (5)

descriptor. It represents an image sequence that has been interconnected into data volume. The volume is subdivides in to equally spaced overlapping blocks and information within each block is represented by histogram of oriented gradients. The local classifier allows finding the probabilities of occlusion. This help to filter the contribution of occlusion from background motion. Finally combine all the action and result is obtained based on the values.

In ref (3) Mohamed et.al proposed 2DPCA. This method is a parallel structure where each path is considered to be an independent action recognition system. First step of this algorithm is human detection technique, which is used to extract the clear silhouettes from the image. Then frame alignment technique is using for aligning the object in centre of every frame. Ref (2) proposes a method of shape matching based on feature and brightness based method. Shape matching method is based on the shape of the object while brightness based method is based on the brightness of the object. Brightness based information is obtained from the correspondence values. This makes direct use of pixel brightness values. It uses the gray values within the visible portion of the object. This method contain two steps, first step find the correspondence of the object. This method is based on the learning algorithm. In the second approach classifiers are built from the correspondence values. Here object is considered as a point set and shape of the object is considered as the subset of point set. In more practical cases the shape is considered as the discrete subset of its points from contour.

IV. PROPOSED WORK

This proposed method of human action recognition uses a number of correlated poses. The extracted normalized silhouettes are used as input features for the BoCP model. Bag of correlated poses are using for encoding the local features of action. Which take many advantages over normal bag of visual words (BoVW) (11). In the traditional bag of visual models each feature vector has assigned to its closest codeword. After the descriptor extraction fuzzy C-means clustering algorithm is using for generating the codebook. To reduce the dimensionality of computed feature Principal Component Analysis and Linear Discriminant Analysis are using.

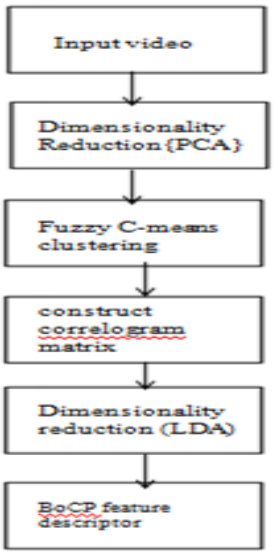

Fig. 2: Block Diagram

A. *Dimensionality Reduction:*

Both the original silhouette feature and final correlogram representation are in very high dimensionality. Due to the curse of dimensionality it is impractical to use the original long feature vectors for classification. Therefore the use of a dimensionality reduction method is necessary. There are two stages where dimensionality reduction is needed in this project. The first stage is before feeding silhouette feature vector in to the fuzzy C-means clustering algorithm, and the second stage is for the reduction of the final correlogram representations. Here unsupervised PCA is using at the first stage and a combination of PCA and LDA are using at the second stage.

1) Principal Component Analysis (PCA):

PCA is a statistical method that uses orthogonal transform to convert a set of correlated values into set of uncorrelated values called principal component (9). The number of principal component is less than or equal to the number of original variables. Here the first principal component has the largest possible variance and the remaining component have comparatively less variance.

2) Linear Discriminant Analysis:

Linear Discriminant Analysis is used in pattern recognition to find linear combination of features which separate two more classes of objects or events. The resulting combination can be used in linear classifier. In LDA one dependent variable can be represented as the linear combination of other features. Here the dependent variable is assumed to be normally distributed. This is the fundamental assumption of LDA. Normally LDA is closely related to PCA. KL transformation is another method using for discriminant analysis.

B. *C-Means Clustering:*

In fuzzy clustering data elements can belong to more than one cluster. These indicate the strength between the data elements and particular cluster. Fuzzy clustering is also called soft clustering. One of the most widely used fuzzy clustering algorithms is Fuzzy C-means clustering (FCM) (7). The FCM algorithm portion the finite collection of n elements in to collection of C fuzzy clusters with respect to some given criteria. In this algorithm a member ship function is assigned to each data point, which is on the basis of the cluster centre and the shortest distance between the cluster centre and data point (8). The summation of each membership data point is equal to one. The concept of fuzzy partition is important in cluster analysis and for identification technique that are based on fuzzy clustering.

C. *Correlogram of Poses:*

The current BoVW model discards all the geometric information, therefore the original BoVW model is extend to BoCP that take the advantages of the local and global features. A color correlogram indicates the occurrence of two colors that are at certain distance from each other. Thus a correlogram have more structural information than a flat histogram. Each element in BoCP denotes the occurrence of two body poses taking place at a certain time difference from each other.

V. EXPERIMENTAL RESULT ANALYSIS

Training of video consists of a number of steps. The action to be tested is given as input video. After that this video is dividing into a sequence of frames. Then each frame is trained by a number of processes. Training of video consists of a number of steps. The action to be tested is given as input video. After that this video is dividing into a sequence of frames. Then each frame is trained by a number of processes. The video to be tested is giving as input and features are extracted. Codebook is created by using these features. Then these features are clustered by fuzzy C-means clustering. The clustering algorithm minimizes the variance between the clusters and data. And performed action is recognized based on the feature value.

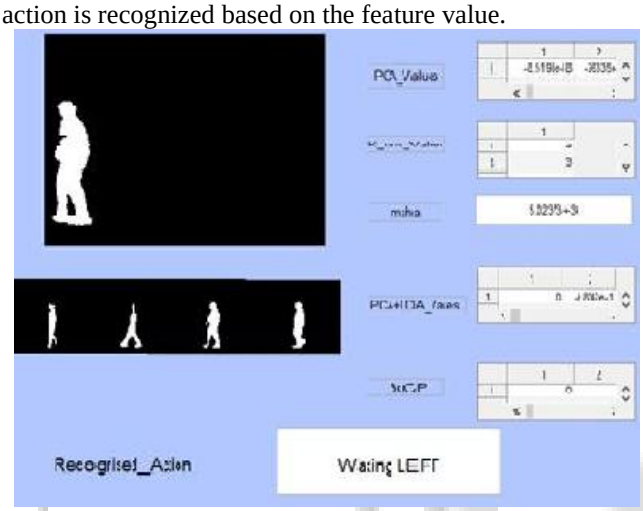

Fig. 3: Testing Of Video

VI. CONCLUSION AND FUTURE SCOPE

This project proposes a new representation namely BoCP (Bag of Correlated Poses) for action recognition. In BoCP model the unique way of considering temporal structural correlation between consecutive human poses encoded more information than the traditional Bag of feature models. This method utilizes the unsupervised method like PCA and LDA for dimensionality reduction. Also this takes the advantages of fuzzy C-means clustering for the clustering purpose. This takes the soft-assignment strategy to preserve the visual word ambiguity. With more feature descriptors and advanced dimensionality reduction methods performance of the system can be increased.

REFERENCE

[1] X. Cao, B. Ning, P.Yen and X. Li, "Selecting key poses on manifold for pair wise action recognition", IEEE Trans. Indust. Infor. Vol.8. no.1.pp. 168-177. Feb. 2012.
[2] D.Weinland, M. Ozuysal and P. Fua, 'making action recognition robust to occlusion and view point changes" in proc. ECCV, 2010, pp. 635-648
[3] Mohamed A Naiel, Moataz M Abdelwahab, " multiview human action recognition system employing 2DPCA" in proc oct. 2000
[4] L.k.Saul and S.T.Roweis, "Think globally, fit locally, unsupervised learning of low dimensional mmanifold" J.Mach.Learn. Res, vol 4, pp.119-155, 2003.
[5] L.Shao and X.Chen, "Histogram of body poses and spectral regression discriminant analysis for human action categorization" in Proc. BMVC, 2010, pp.88.1-88.11
[6] T. Goodhart, P. yan and M. Shah, "Action recognition using spatiotemporal regularity based feature" in proc. IEEE Coput. Soc. Conf. CVPR, vol. 2.Jun.2005, pp. 524-531
[7] S.Chattopadhyay, D.K Pratihar, S.D Sarkar "A comparative study of fuzzy C- means clustering and entropy based fuzzy clustering algorithm" computing and informatics vol 30, 2011, pp 701-720.
[8] Albayrak.S, "fuzzy C-means clustering on medical diagnostic system" proc. Int. XII Turkish symp.on artif. Intel. NN 2003
[9] L.K Saul and S.T Roweis, "think globally fit locally: unsupervised learning of low dimensional manifolds" J.Mach. Learn. Res. Vol.4, pp.119-155, 2003.
[10] L.Shao, D.Wu and X. Chen "Action recognition using correlogram of body poses and spectral regression" in Proc. Int. Conf. Image process. 2011 pp 209- 212.
[11] P. Dollar, V. Rabuad, G. Cotterell and S. Belongie "Behavior recognition via sparse spatio-temporal features" in proc. 2nd joint IEEE int. workshop Vis. Surveillance performance Oct. 2005, pp 65-72.
[12] X.Sun, M. Chen and A. Hauptmann, "Action recognition via local descriptor and holistic features" in Proc. IEEE Comput. Soc. Conf. CVPR Workshop, Jun. 2009, PP 58-65.

