International Journal of Engineering Research-Online
A Peer Reviewed International Journal
Articles available online http://www.ijoer.in

Vol.3., Issue.1, 2015

REVIEW ARTICLE

ISSN: 2321-7758

ACTION RECOGNITION BY ANALYSING HUMAN SILHOUETTE VIDEO – A SURVEY

DIVYA RAJAN N T[1*], Mr.BLESSINGH T S[2], Dr.SREEJA MOLE S.S[3],

[1]PG Student, ECE Department, Narayanaguru College of Engineering, Manjalumood, Tamil Nadu, KK District, India

[2]Assistant Professor, ECE Department, Narayanaguru College of Engineering, Manjalumoodu, KK District, Tamil Nadu

[3]Head of the Department ECE, Narayanaguru College of Engineering, Manjalumoodu, Tamil Nadu, KK District, India

Article Received: 08/01/2015

Article Revised on: 16/01/2015

Article Accepted on:19/01/2015

DIVYA RAJAN N T

ABSTRACT

This paper proposes a novel method for human action recognition. It has wide applications including access control in special areas, human identification at a distance, crowd flux statistics, detection of anomalous behaviors etc. An action video can be represents as a number of image frames. Bag of Correlated Poses is introduced to encode the local features of action. This paper review recent developments and general strategies of all the stages. The processing frame work of visual surveillance in dynamic scenes include the following stages: modeling of environments, detection of motion, classification of moving objects, tracking, understanding and description of behaviors, human identification and fusion of data from multiple cameras. The proposed scheme has the advantages of local and global features and provides discriminative representation for human action. To utilize the property of visual word ambiguity, it adopts the soft assignment strategy to reduce the computational complexity and quantization error.

Key words- Bag of Correlated poses, quantization error, soft assignment scheme, visual surveillance.

INTRODUCTION

In recent years, human action recognition has drawn in increasing attention of researchers, due to its growing applications in areas, such as video surveillance, robotics, human- computer interaction, user interface design, and multimedia video retrieval. Human motion analysis is currently one of the most active research topics in computer vision. Action recognition is still a challenging problem due to difficulties such as intra class variation, scaling, Occlusion and cluttering. Motion and structure are the main cues of the action occurring in a video sequence unfortunately, most of human motion analysis methods are constrained with the assumption of view dependence, i.e., actors have to face a camera or have to be parallel to a viewing plane.

Most of the current methods for 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, sometimes impossible, to achieve in realistic scenarios. One main difficulty lies in the fact that the motion field in the action region is contaminated by the background motions. The goal of this paper is to recognize human action in dynamic and crowded environments by the use of the action models that are the trained on data with clean background. To achieve the goal, we need to develop techniques to obtain stable and generalize features and feature descriptors that are able to characterize action but incentive to cluttered background motion. Motion information includes the body or body part movements and position translations, while structure information covers body poses, there occurring orders and relative positions. To characterize an action we need to effectively encode both of them to obtain an informative representation.

Due to the limitations of view-dependent characteristic a large number of human motion analysis methods 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 a research progress on removal of the effect on human motion analysis methods had been reported in recent years. Hence it is timely comprehensively review recent research on view invariant human motion analysis.

2D PRINCIPAL COMPONENT ANALYSIS

Two dimensional PCA (Principal Component Analysis) techniques [4] for facial recognition have many advantages over the PCA method. This method consists of a parallel structure where each path can be considered as an independent human action recognition system that processes every frame. First step is human detection technique, which is used to extract clear silhouettes for people. Then a frame alignment technique is applied to align the object in the centre of every frame. The MEI or MHI are used to generate different patterns depending on the input alignment silhouettes a suitable transform, i.e. 2D-DCT, can be used to compress the generated patterns from the MEI or the MHI stages. The 2D PCA algorithm is applied in both training and testing phase for feature extraction from the input patterns in the spatial domain or the transform domain. The K nearest neighbor (KNN) classifier is used to find the most likely class. Finally, a majority voting technique is used to decide the corresponding action based on the output of multiple classifiers. The main advantages of the algorithm are it is simpler for image feature extraction, better recognition rate and more efficient in computation. However, it is not as efficient as PCA in terms of storage requirements.

MANIFOLD FOR PAIR WISE ACTION RECOGNITION

This paper [1] proposes a novel approach for key pose selection, which models the descriptors space utilizing a manifold learning technique to recover the geometric structure of the descriptors on a lower dimensional manifold. One of the descriptor have been extracted a compact set called codebook needs to be learned. Thus an action can be represented as a collection of words from the codebook. In order to catch the relationship between the high- dimensional descriptors and find typical key pose a refined model for the future space is needed. For this purpose, we introduce manifold learning algorithm [5], which learns an internal model of the input data and the projects the high dimensional descriptors onto the low dimensional manifold. Page rank-based centrality measure is developed to select key poses according to the recovered geometric structure. Page rank [6], which assigns relative scores to all nodes in the graph based on the recursive principle. Initially each node is assigned an equal relevance score 1. Then iteration process is executed in which the relevance of the node is determined recursively based on a Markov chain model. The page rank measure adapted from web search engine tends to assigns convergence score to the nodes in the central area. The adjacent nodes provide relevance score for each other in each circulation so the nodes adjacent to central nodes also get high score. On codebook has been obtained an action can be represented by a collections of key poses in it. Every frame of the action video corresponds to a descriptor, which can be matched to the nearest neighbor in the codebook. Thus, an action translated to series of words in the codebook. The frequency of the words in the series can be accumulated into a histogram to represent the action and be used for recognition. In the testing process the pose descriptors are first extracted from the testing video. Since each key pose in the obtained codebook corresponds to an original pose descriptors of the testing video can be compared with the key poses without LLE mapping. Therefore the action video can be directly translates to the histogram according to the codebook. SVM makes the final prediction by using this histogram as input feature vector.

3D HISTOGRAM OF ORIENT GRADIENTS (HOG)

The 3D histogram of oriented gradient (HOG) descriptor [2] used represent sequence of images that have been concatenated into a data volume. It relies on the 3D extension of the HOG descriptor [7] represents image sequence that have been concatenated into a data volume. The volume is subdivides into equally spaced overlapping blocks and information within each block is represented by a histogram of oriented 3D spatial-temporal gradients [8]. The resulting block descriptions are embedded temporally at each special location, providing a discriminative representation that has fixed dimension independent of the duration of a sequence and hence can be easily fed to a classifier. By contrast to HOG and BoW, the feature descriptors are not spatially integrated into a global representation, i.e. by concatenating the blocks into single vector (HOG) or by computing a location independent histogram of the blocks (Bow). Preserving location dependent information introduces additional discriminative power. Moreover, the local classifier let us also estimate probabilities for occlusion, which we use to filter out contribution from cluttered and occluded region when finally combining the local action assignment into a global decision. Computing the descriptors involves the following steps: first, the region to be characterized is portioned into regular cells and histogram h of 3D gradient orientations is computed in each one. This compactly represents temporal and spatial texture information and is invariant to local deformation. Histogram for all cells in a small neighborhood are then concatenated in to a block descriptor B to which SWIFT- like L2 normalization with clipping is applied to increase robustness. Since the block overlap with each other this yields redundant representation which increases discriminative power because normalization emphasizes different bins in different blocks.

OBJECT TRACKING

Region based tracking

Region based tracking algorithm track objects according to variations of image regions corresponding to the moving objects. For these algorithms the background image is maintained dynamically, and motion region are usually detected by subtracting the background from the current image. Recently in [9] proposes an adaptive background subtraction method in which color and gradient information are combined to cope with shadows and unreliable color cues in motion segmentation. Tracking is then performed at three levels of abstraction: regions, people and groups. Each region has a bounding box and regions can merge and split. A human image is composed of one or more regions grouped together under the condition of geometric structure constraints on the human body, and human group consist of one or more people grouped together. Therefore, using the region tracker and the individual color appearance model perfect tracking of multiple people is achieved even during occlusion.

Feature Based Tracking

Feature based tracking algorithm perform recognition and tracking of objects by extracting elements, clustering them into higher level features and then matching the features between images. Feature based tracking algorithm can further be classified into three subcategories according to the nature of the selected feature: global feature based algorithm, local feature based algorithm, and dependence based algorithm. The features used in global feature based algorithm include centroids, perimeters, areas, some orders of quadratures and colors. Even when occlusion happens between two persons during tracking as long as the velocity of the centroids can be distinguished effectively tracking is still successful. Using motion estimation based on Kalman filter the tracking of a non rigid moving object is successfully performed by minimizing a feature energy function during the matching process.

Model Based Tracking

Model based tracking algorithms track objects by matching projected object models, produced with prior knowledge to image data. The models are usually constructed off line with manual measurement like CAD tools or computer vision techniques. As model based rigid object tracking and model based non rigid object tracking are quite different review separately model based human body tracking (non rigid object tracking) and model based vehicle tracking (rigid object tracking). The general approach for model based human body tracing is known as analysis by synthesis and it is used in a predict match update style. First the pose of the model for the next frame is predicted according to prior knowledge and tracking history. Then the predicted model is synthesized and projected into the image plane for comparison with the image data. A specific pose evaluation function is needed to measure the similarity between the projected model and the image data. According to different search strategies, this is done either recursively or using sampling techniques until the correct pose is finally found and is used to update the model. Pose estimation in the first frame needs to be handled specially. Generally model based human body tracking involves three main issues, first the recognition rate of object based on 2-D image feature is low because of the nonlinear distortion during perspective projection and the image variations with the view point's movement, second these algorithms are generally unable to recover 3-D pose of object and the third is the stability of dealing effectively with occlusion, overlapping and interference of unrelated structures is generally poor.

PROPOSED METHOD

The proposed method uses a novel method to human action recognition that uses a number of correlated poses. Each silhouette video is converting into a set of image frames for feature extraction. The extracted normalized silhouettes are used as the input features for the BoCP model. Bag of correlated poses (BoCP) are using for using for encoding the local feature of action. Which take many advantages over normal bag of visual words (BoVW). Bag of correlated poses has advantages of both local and global representation. In the traditional BoVW model each feature vector can be assigned to its closest codeword based on the Euclidean distance. After the descriptor extraction K means clustering is using to generate the codebook. Clustering is the process of partitioning the data set in to subsets (clusters), so that the data in each subset shares some common trait – often 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 result prove the viability of the complementary properties of two descriptors and the proposed approach outperforms of the complementary properties of two descriptors and the proposed approach outperforms the state of the art methods on the IXMAS action recognition dataset.

CONCLUSION

This paper utilized a soft-assignment strategy to preserve the visual word ambiguity that was usually disregarded during the quantization process. Here presented an overview of recent developments in visual surveillance within a general processing framework for visual surveillance systems. Here three techniques for motion segmentation are addressed: background subtraction, temporal differencing and optical flow. This paper also studied different approaches to tracking. It proposed a method for action recognition by analyzing the two dimensional principal component. At the end of this survey a new method is proposed. It includes some detailed discussion on future direction such as occlusion handling, fusion of 2D tracking and 3D tracking. As to fusion of data from multiple cameras, it reviewed installation, object matching, switching of data and data fusion. The proposed system utilized a soft assignment strategy to preserve the visual word ambiguity that was usually disregarded during the quantization process after K means clustering.

REFFERENCES

[1]. X. Cao, B. Ning, P.Yen, and X. Li, "Selecting key poses on manifold for pair wise action recognition", IEEE Trans. Indust. Inform. 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]. Weiming Hu, Tieniu Tan, "Survey on visual surveillance of object motion and behaviors", IEEE Trans. System. Vol.v34, no. 3, Aug. 2004.

[4]. Mohamed A Naiel, Moataz M Abdelwahab, "multi view human action recognition system employing 2DPCA" in proc. Oct. 2000.

[5]. 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.

[6]. S.Brin and L.Page, "The anatomy of a large scale hyper-textual web search engine", Compt. Netw. ISDN Syst. Vol. 30, pp. 107-117, 1998.

[7]. Dalal, N. Triggs, B.: Histogram of oriented gradients for human detection. In: CVPR (2005).

[8]. Klaser, A. Marszalek, M. Schmid: a spatio-temporal descriptor based on 3D gradients. In: BMVC (2008).

[9]. S. Mckenna, S. Jabri, Z. Duric, A.Rosenfeld and H.Wechsler "tracking groups of people." Comput. Vis. Image understanding. Vol. 80, no. 1, pp. 42-56, 2000.

[10]. M. Blank, L. Glorelick, E. Shechtman, M. Irani, and R. Basri, " action as space time shapes" in proc. 10th IEEE ICCV, vol. 2, Oct. 2005, pp. 1395-1402.

[11]. L. Fei-Fei, P. Perona, "a Bayesian hierarchical model for learning natural scene categories" in proc. IEEE ICASSP, Mar.-Apr. 2008, pp. 745-748.

[12]. T. Goodhart, P. Yan and M. Shah, "Action recognition using spatiotemporal regularity based features," in proc. IEEE Coput. Soc. Conf. CvPR, vol. 2.Jun. 2005, pp. 524-531.

[13]. B. Leibe, A. leonardis, and B. Schiele, "combined object categorization and segmentation with an implicit shape model," in proc. Workshop statist. Learn. Comput. Vision (ECCV), 2004, pp. 17-32

[14]. 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

[15]. X.Sun, M. Chen and A. Hauptmann," Action recognition via local descriptor and holistic features," in proc. IEEE Comput. Soc. Conf. CVPR Workshops, Jun. 2009, pp. 58-65.

[16]. [16] M.Varma and B. Babu, "more generality in efficient multiple kernel learning," in proc. 26th annu. Int. Conf. march. Learn., 2009, pp. 1065-1072