A STUDY ON HUMAN ACTIVITIES RECOGNISATION SYSTEM USING KNN CLASSIFICATION

Amandeep Kaur, Er. Jaspreet Kaur

Abstract— **The Research work Human action recognisation region of interest part of video with hybrid technique and enhancement of Non –region of interest part is studied in this work. In recent years, automatic human activity recognition has drawn much attention in the field of video analysis technology due to the growing demands from many applications, such as surveillance environments, entertainment environments and healthcare systems. There is problem is the sparse decoding data loss problem due to ROI and NOI-ROI region of the action detected video. To enhance the human action detection part of the video and non auctioned part of the video with the help of Hybrid technique. The SVM is used detect the 2D/3D pose of the human and the accuracy of the system on different datasets are tested.**

Index Terms—**2D/3D Pose, ROI, Non-ROI etc.**

I. INTRODUCTION

Human action recognition has been an attractive and popular research topic in recent two decades. Most previous works in this topic employed a frame-by-frame comparison to trained action models for classifying a newly arrived video sequence, which is computationally expensive due to the following facts:

- The consecutive frames in a video are correlated/ similar in temporal domain; hence it is redundant to compare every frame for classification.
- In some cases, only a few frames in a video are sufficient for discrimination of basic actions [1].

Human action recognition has a wide range of applications such as video content analysis, activity surveillance, and human-computer interaction [1]. As one of the most active topics in computer vision, much work on human action recognition has been reported [2]. In most of the traditional approaches for human action recognition, action models are typically constructed from patterns of low level features such as appearance patterns, optical flow [1], space-time templates, 2D shape matching, trajectory-based representation and bag-
Of-visual-words (BoVW) . However, these features can hardly characterize rich semantic structure in actions. Inspired by recent development in object classification
[3], we introduce a high-level concept named "action unit" to describe human actions, as illustrated in Figure 1.1.For example, the "golf-swinging" action contains some representative motions, such as "arm swing" and "torso twist". They are hardly described by the low-level features mentioned above. On the other hand, some correlated space-time interest points, when combined together, can characterize a representative motion. Moreover, the key frame is important to describe an action; and a key frame may be characterized by the co-occurrence of space-time interest points extracted from the frame. The representative motions and key frames both reflect some action units, which can then be used to represent action classes. we propose using high-level action units for human actions representation. Typically, from an input human action video, hundreds of interest points are first extracted and then agglomerated into tens of action units, which then compactly represent the video. Such a representation is more discriminative than traditional BoVW model.

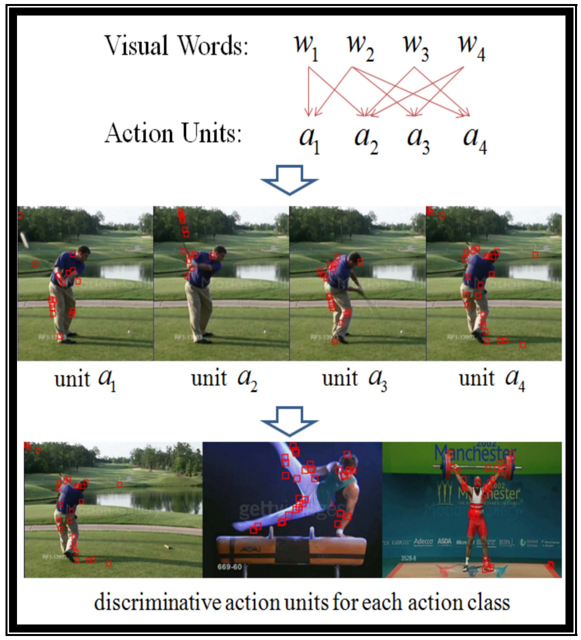
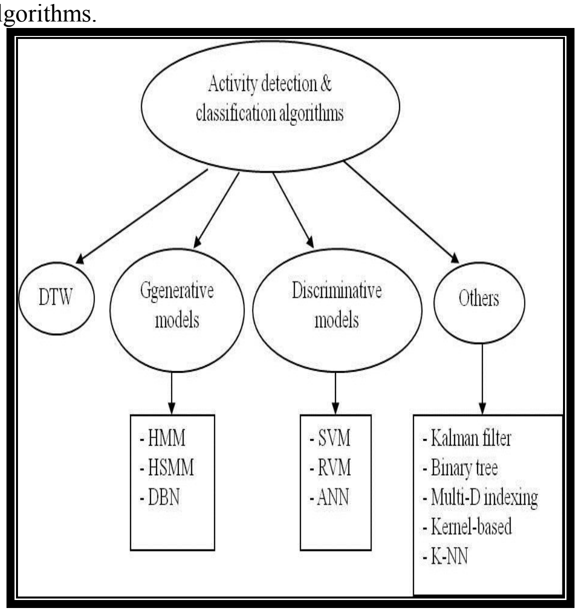


Figure 1.1: A single interest point may have multiple meanings in different contexts. Some correlated interest points together can construct an action unit which is more descriptive and discriminative. A video sequence can be represented by a few action units, and each action class has its own representative action units.

In recent years, automatic human activity recognition has drawn much attention in the field of video analysis technology due to the growing demands from many applications, such as surveillance environments, entertainment environments and healthcare systems. In a surveillance environment, the automatic detection of abnormal activities can be used to alert the related authority of potential criminal or dangerous behaviors, such as automatic reporting of a person with a bag loitering at an airport or station. Similarly, in an entertainment

Manuscript received April 29, 2015
Amandeep Kaur, M.tech Scholar Deptt. of CSE, Guru Kashi University, Talwandi Sabo
Er. Jaspreet Kaur, Assistant Professor Deptt. of CSE, Guru Kashi University, Talwandi Sabo

environment, the activity recognition can improve the human computer interaction (HCI), such as the automatic recognition of different player's actions during a tennis game so as to create an avatar in the computer to play tennis for the player. Furthermore, in a healthcare system, the activity recognition can help the rehabilitation of patients, such as the automatic recognition of patient's action to facilitate the rehabilitation processes. There have been numerous research efforts reported for various applications based on human activity recognition, more specifically, home abnormal activity [1], ballet activity [2], tennis activity, soccer activity, human gestures , sport activity , human interaction, pedestrian traffic [1] and simple actions , and healthcare applications.
The categories for activity detection and classification algorithms.



The dynamic time warping (DTW) , a method for measuring similarity between two temporal sequences, which may vary in time or speed, is one of the most common temporal classification algorithms due to its simplicity; however, DTW is not appropriate for a large number of classes with many variations. Some probability-based methods by generative models (dynamic classifiers) are proposed such as Hidden Markov Models (HMM) [2] and Dynamic Bayesian Networks (DBN) . On the other hand, discriminative models (static classifiers) such as Support Vector Machine (SVM), Relevant Vector Machine (RVM) and Artificial Neural Network (ANN) , can also be used in this stage[3]. In addition to the dynamic and static classifier difference nature, another main difference between generative models and discriminative models [4] is that the generative classifiers commonly learn a model of the joint probability, $p(x,y)$, of the input x and the label y, or equivalently the likelihood $p(x|y)$ according to Bayes' rule; while the discriminative classifiers model the posterior $p(y|x)$ directly. Therefore, the generative models can be used to simulate values of any variables in the models, while the discriminative models allow only sampling of the target variables conditional on the observed variables. For both of the probability model-based algorithms, including generative models and discriminative models, their performance relies on extensive training dataset. Therefore, other methods are proposed, such as Kalman filter [5], binary tree [4], multidimensional indexing [1], and K nearest neighbor (K-NN) [2]. Different classification algorithms usually require different sets of suitable feature representations.

1.1 Static Camera

In static camera segmentation, the camera is fixed in a specific position and angle. Since the background never moves, it is natural to build a background model in advance, so that the foreground object can be segmented from the image of the background model.

1.2 Background Subtraction

The most common method for static camera segmentation is background subtraction due to its simplicity and efficiency. The background model contains only the stationary background scene without any foreground object, and any image change is assumed to be caused only by moving objects. Hence the foreground object can be obtained by subtracting the current image of the background image, followed by a magnitude thresholding to obtain the segmentation mask. The segmentation mask often contains rough and fractional foreground object(s) and usually requires some post-processing, such as closing and opening morphological operations. The background subtraction has been extensively applied in all kinds of scenarios with various improved modifications. For example, for real-time human body tracking [3], the color distribution of each pixel in the background is first modeled with a Gaussian with a full covariance matrix. This background scene texture map is considered to be class zero. The foreground textures in different classes are grouped by the mean of a point and the covariance associated with that point. Another improvement is to discriminate moving objects, ghosts and shadow [4], based on statistical assumptions, with object-level knowledge, of moving objects, apparent objects (ghosts) and shadows. Besides, in order to overcome the limitation of the background subtraction on stationary background.

II. RELATED WORK

This chapter includes the literature of the research work that I have to implement in my research work. The different research papers of different researchers are studied that is given below:

Haoran Wang et.al [2014] have proposed using high-level action units to represent human actions in videos and, based on such units, a novel sparse model is developed for human action recognition. There are three interconnected components in our approach. First, we propose a new context-aware spatial temporal descriptor, named locally weighted word context, to improve the discriminability of the traditionally used local spatial-temporal descriptors. Second, from the statistics of the context-aware descriptors, we learn action units using the graph regularized nonnegative matrix factorization, which leads to a part-based representation and encodes the geometrical information. These units effectively bridge the semantic gap in action recognition. Third, we propose a sparse model based on a joint $l2,1$-norm to preserve the representative items and suppress noise in the action units. Intuitively, when learning the dictionary for action representation, the sparse model captures the fact that actions from the same class share similar units. The proposed approach is evaluated on several publicly available data sets.[1]

Raviteja Vemulapalli et.al [2014] has proposed recently introduced cost-effective depth sensors coupled with the real-time skeleton estimation algorithm have generated a renewed interest in skeleton-based human action recognition. Most of the existing skeleton-based approaches use either the joint locations or the joint angles to represent a human skeleton. In this paper, we propose a new skeletal representation that explicitly models the 3D geometric relationships between various body parts using rotations and translations in 3D space. Since 3D rigid body motions are members of the special Euclidean group SE(3), the proposed skeletal representation lies in the Lie group SE(3)X………X SE(3), which is a curved manifold.[2]

Muhammad Shahzad Cheema et. al [2014] have studied methods are distinguished by naive but efficient feature extraction, sparse coding, instance-based learning and latent factor analysis. In particular, we (a) devise an efficient discriminative key poses approach for action recognition in videos that is independent of temporal context (b) present an efficient and scalable nearest affine hull method to HDLSS activity classification based on least squares optimization and QR-factorization (c) present a hierarchical bilinear factorization approach of style and content separation to recognize actions and actors in 3D data (depth, motion capture, motion history volumes) and (d) propose a non-negative matrix factorization based approach to determine action signatures from videos that are later used as saliency maps for classification of images.[3]

Shian-Ru Ke et. al [2013] have surveys extensively the current progresses made toward video-based human activity recognition. Three aspects for human activity recognition are addressed including core technology, human activity recognition systems, and applications from low-level to high-level representation. In the core technology, three critical processing stages are thoroughly discussed mainly: human object segmentation, feature extraction and representation, activity detection and classification algorithms. In the human activity recognition systems, three main types are mentioned, including single person activity recognition, multiple people interaction and crowd behavior, and abnormal activity recognition. Finally the domains of applications are discussed in detail, specifically, on surveillance environments, entertainment environments and healthcare systems. Our survey, which aims to provide a comprehensive state-of-the-art review of the field, also addresses several challenges associated with these systems and applications.[4]

III. PROBLEM DEFINITION

In the human action recognized research work different problems are studied from the review of different researchers. Their are different problems that the previous work is only for 2D/3D pose estimation of the human body modeling. Another human activity of great interest to many researchers due to the fact that the loss of ability to walk correctly can be caused by a serious health problem, such as pain, injury, paralysis, muscle damage, or even mental problems. In the action recognized system It is Difficult to identify the side view of the person with some cameras, we can only identify the front and back side of the in a video. Another problem is that there is sparse decoding data loss problem due to ROI and NOI-ROI region of the action detected video.

IV. OBJECTIVE

There are following objectives that we have to fulfill in this research work that are given below:

- To enhance the human action detection part of the video and non actioned part of the video with the help of Hybrid technique.
- To resolve the problem of 2D/3D pose problem with the help of SVM classifiers.
- Analyse the result being obtained for the existing techniques.
- Calculate the accuracy of the system on different datasets.

V. METHODOLOGY

This research work is to implement the theft security system based on face reorganization. It is based upon GUI (graphical user interface) in MATLAB. It is an effort to further grasp the fundamentals of MATLAB and validate it as a powerful application tool. There are basically different files. Each of them consists of m-file and figure file. These are the programmable files containing the information about the images. We proposed a framework for human action detection in a video.

The video data set that we have to test and train and find the region of interest and Non-ROI part of the video and after that process the ROI part to detect the action of the human with SVM and K-NN classification and enhance the Non –ROI part of the video. Find the accuracy of the detected part.

By estimating the region of interest in video with ROI and Non-ROI part of the video. The proposed work under the following Steps:

Step 1: Acquire the input video data set that is to be processed.

Step 2: Insert point detection to detect the video.

Step 3: Apply the LWWC descriptor to get the in formation of the video.

Step 4: After Step:3 also apply the GNMF based action unit & also apply the action unit based representation to represent action and get the ROI and Non-ROI part of the video .

Step 5: Recognize the human action from the ROI part with hybrid technique and enhance the Non-ROI part.

Step 6: Repeat the step 1 to step 3 for test data set of the video

Step 7: After Ste p 5: label the recognized action with the help of action label.

Step 8: Stop

CONCLUSION

In a surveillance environment, the automatic detection of abnormal activities can be used to alert the related authority of potential criminal or dangerous behaviors, such as automatic reporting of a person with a bag loitering at an airport or station. There are different problems that the previous work is only for 2D/3D pose estimation of the human body modeling. Another human activity of great interest to many researchers due to the fact that the loss of ability to walk correctly can be caused by a serious health problem, such as pain, injury, paralysis, muscle damage, or even mental problems. In the future work KNN and SVM is implemented to detect the human actions.

REFERENCES

[1]. Haoran Wang et.al "Action Recognition Using Nonnegative Action Component Representation and Sparse Basis Selection" IEEE transactions on image processing, vol. 23, no. 2, February 2014.

[2]. Raviteja Vemulapalli et.al "Human Action Recognition by Representing 3D Skeletons as Points in a Lie Group" IEEE -2014.

[3]. Muhammad Shahzad Cheema et. al "Efficient Human Activity Recognition in Large Image and Video Databases"2014.

[4]. Shian-Ru Ke et. al " A Review on Video-Based Human Activity Recognition" computers ,ISSN 2073-431X, www.mdpi.com/journal/computers Computers 2013.

[5]. Duong et.al " Activity Recognition and Abnormality Detection with the Switching Hidden Semi-Markov Model" . In Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR), San Diego, CA, USA, 20–25 June 2005.

[6]. Gorelick et.al " Actions as Space-time Shapes"In Proceedings of the Tenth IEEE International Conference on Computer Vision (ICCV), Beijing, China, 17–21 October 2005.

[7]. Sukthankar et.al "Spatio-temporal Shape and Flow Correlation for Action Recognition"In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Minneapolis, MN, USA, 17–22 June 2007.

[8]. Yamato et.al " Recognizing Human Action in Time-sequential Images using Hidden Markov Model" In Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR), Champaign, IL, USA, 15–18 June 1992.

[9]. W. Brendel et. al. "Activities as time series of human postures," in Proc. ECCV, 2010.

[10].Q. V. Le et. al "Learning hierarchical invariant spatio-temporal features for action recognition with independent subspace analysis," in Proc. IEEE Int. Conf. Comput. Vis. Pattern Recognit., Jun. 2011