© INNOVATIVE RESEARCH THOUGHTS | Refereed | Peer Reviewed | Indexed
ISSN : 2454 – 308X | Volume : 07 , Issue : 02 | April - June 2021

Human Activities Recognisation System Using Knn Classification

** Amandeep Kaur*
M.tech IT-Research Scholar
Guru Kashi University
Talwandi Sabo
aman.chouhan70@gmail.com

Er. Jaspreet Kaur
Assistant Professor Deptt. of CSE
Guru Kashi University
Talwandi Sabo
jaspreet11cse@gmail.com

Abstract: Human action acknowledgment is an essential zone of PC vision exploration and applications. The objective of the action acknowledgment is a robotized investigation (or understanding) of progressing occasions and their connection from feature information. Its applications incorporate reconnaissance frameworks, patient observing frameworks, and a mixture of frameworks that include associations in the middle of persons and electronic gadgets, for example, human-PC interfaces. 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. 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 and find the accuracy of the detected part .

Keywords: Actions, SVM ,KNN, Weizmann dataset , accuracy 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.

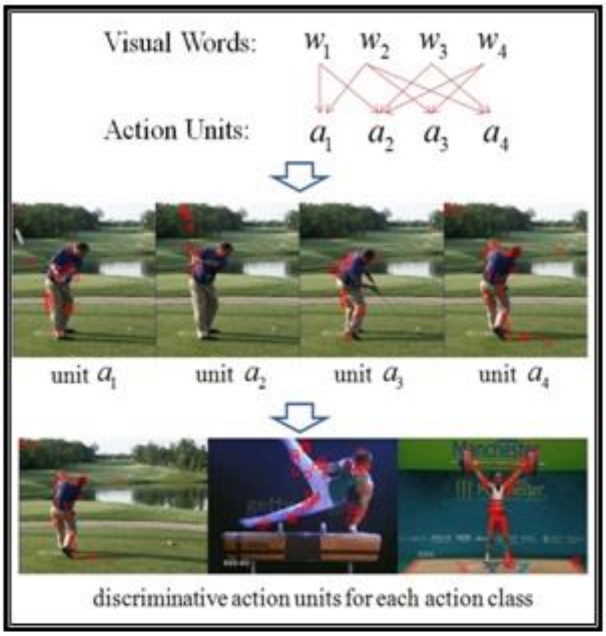


Figure 1.1: A single interest point may have multiple meanings in different contexts. Some correlated interest points together can construct an action unit which is more descriptive and discriminative. A video sequence can be represented by a few action units, and each action class has its own representative action units.
every activity class has its own particular agent activity units. As of late, programmed human action acknowledgment has attracted much consideration the field of feature investigation innovation because of the developing requests from numerous applications, for example, observation situations, excitement situations and health awareness frameworks. In a reconnaissance situation, the programmed discovery of strange exercises can be utilized to caution the related power of potential criminal or hazardous practices, for example, programmed reporting of a man with a pack lingering at an air terminal or station. Also, in an amusement domain, the movement acknowledgment can enhance the human PC connection (HCI, for example, the programmed acknowledgment of distinctive player's activities amid a tennis diversion to make a symbol in the PC to play tennis for the player. Besides, in a human services framework, the action acknowledgment can help the restoration of patients, for example, the programmed acknowledgment of quiet's activity to encourage the recovery forms. There have been various exploration endeavors reported for different applications taking into account human action acknowledgment, all the more particularly, home anomalous movement 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.

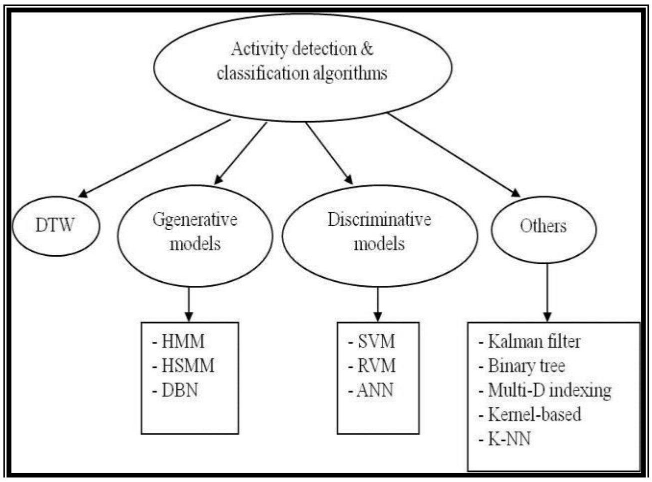


Figure 1.2 : Activity detection and classification

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. 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 awesome enthusiasm to numerous specialists because of the way that the loss of capacity 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.

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.

IV.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:

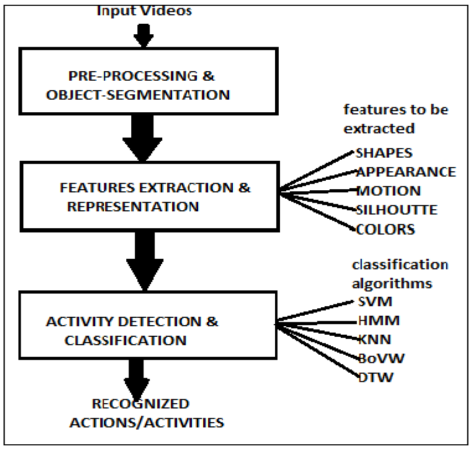


Figure 4.1 –Human Activity Recognition Process

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

Generally speaking, the task of human activity recognition can be divided into three levels comprising of pre-processing and object segmentation, feature extraction and representation and activity detection and classification as shown in figure

Pre-processing & segmentation

The pre-processing stage involves the extraction of frames from the video as most of the previously done work in the field of human activity employs a frame-by-frame processing.

Segmentation is done to extract the target object from the frames depending upon the camera mobility from which the videos were captured. For the static cameras, the camera alignment is fixed in a specific position and angle. As the background never moves, One can use the background subtraction method, wherein the current image of the background image is subtracted to get the required foreground object. On the other hand, contrasting to the simplicity of static camera segmentation, moving camera segmentation is quite challenging due to the fact that both the motion of the target object and the camera orientation and background keeps varying. The most common method for segmenting such videos is identifying the temporal difference between the consecutive frames [6].

Feature Extraction and Representation

Once the region of interest (ROI) is obtained from a frame, feature extraction is done where features like color, silhouette, shape are extracted. In a video sequence, the features that capture the space and time relationship are known as space-time volumes (STV) [2,3,11,51]. The features could be space-time information, body modeling, local descriptors etc [6].

Activity Detection and Classification

Then comes the classification which helps to recognize the human activities on basis of the features extracted. The classifiers use to recognize and classify the actions are SVM, KNN , DTW, HMM etc [6].

V. RESULT

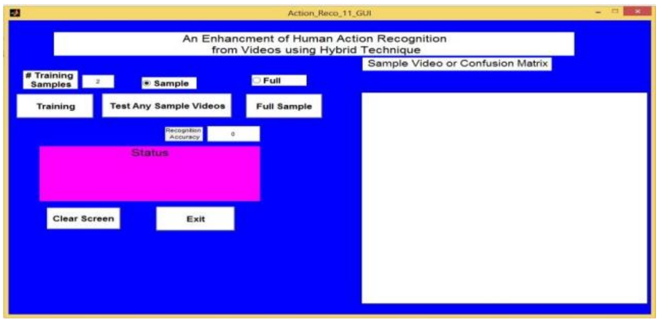


Fig 5.1: Graphical User Interface

The figure 5.1 is the graphical user interface that is used to show the input and the output of the video.In this the video is browsed.

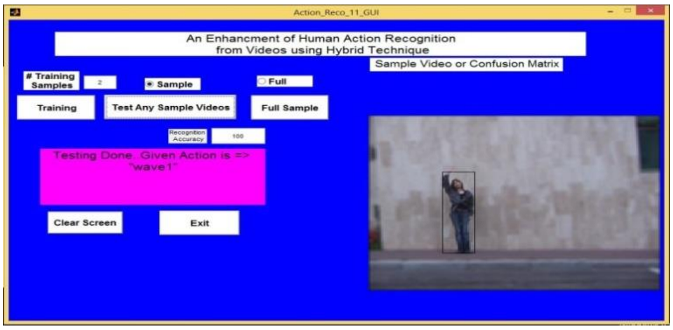


Fig 5.2: Wave1 Action

The Figure 5.2 is displaying the wave1 action from the dataset video I have browsed.

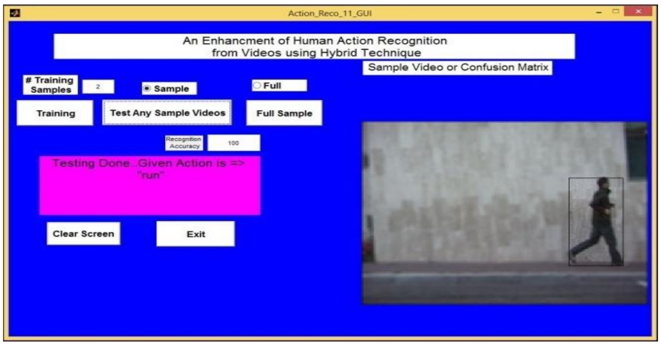


Fig 5.3: Run Action

The Figure 5.3 is displaying the run action from the dataset video I have browsed.

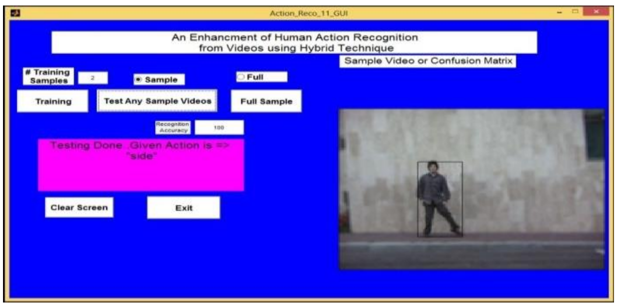


Fig 5.4: Side action

The Figure 5.4 is displaying the side action from the dataset video. I have browsed.

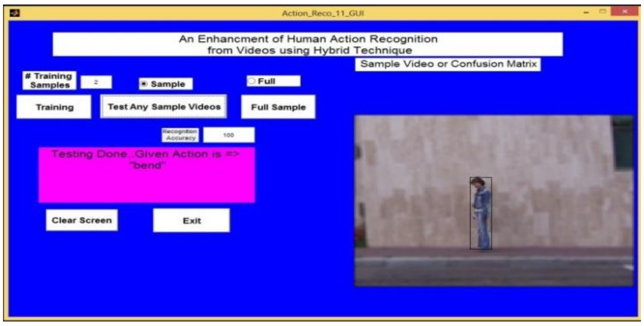


Fig 5.5: Bend action

The Figure 5.5 is displaying the bend action from the dataset video I have browsed.

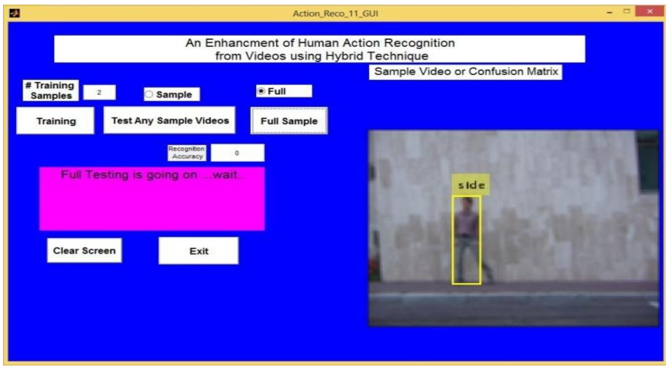


Fig 5.8: Side action

The figure 5.8 is displaying the side action from the dataset video I have browsed.

RESULT FOR MULTIPLE VIDEO

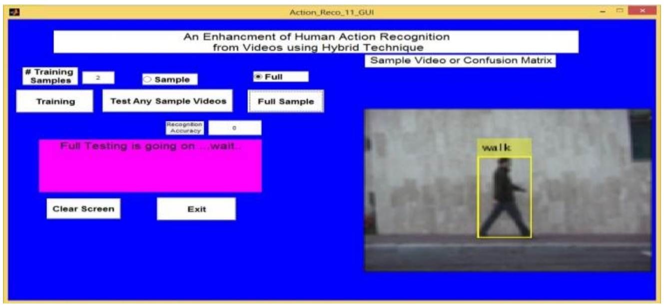


Fig 5.6: Walk action

The Figure 5.6 is displaying the walk action from the dataset video I have browsed.

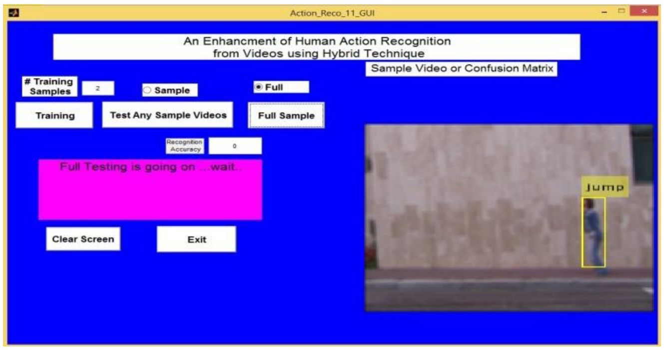


Fig 5.7: Jump Action

The figure 5.7 is dispalying the jump action from the dataset video I have browsed.

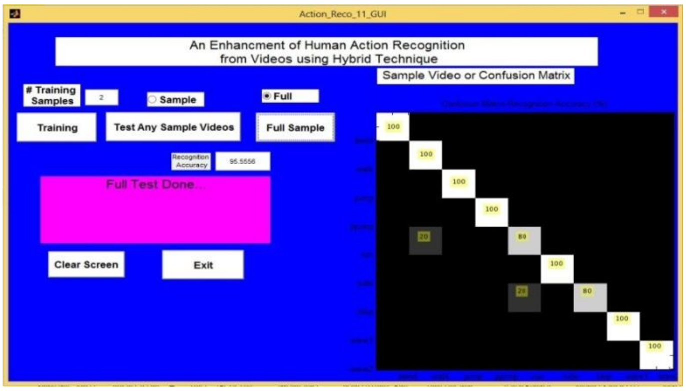
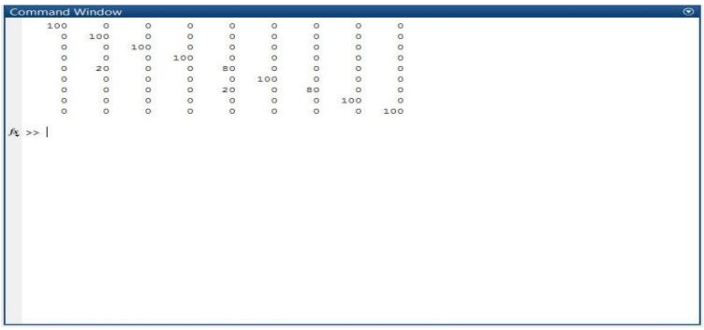


Fig 5.9: Final Result

The Figure 5.9 is displaying the final result with confusion Matrix i have browsed the multiple video.


Command Window

100	0	0	0	0	0	0	0	0
0	100	0	0	0	0	0	0	0
0	0	100	0	0	0	0	0	0
0	0	0	100	0	0	0	0	0
0	20	0	0	80	0	0	0	0
0	0	0	0	0	100	0	0	0
0	0	0	0	20	0	80	0	0
0	0	0	0	0	0	0	100	0
0	0	0	0	0	0	0	0	100

>>

Fig 5.10: Confusion Matrix

Fig 5.10: Table for comparison accuracy of different Researchers

<u>**Accuracy for single and Multiple video**</u>

Table 1: Accuracy table for single video

Fig 5.11: Graph for accuracy table

Approach	Year	Accuracy
Liu et al.	2009	71.2%
Liu et al.	2009	76.1%
Lkizer-cinbis et al.	2010	75.2%
Le et al.	2011	75.8%
Liu et al.	2012	70.4%
Wang et al.	2013	84.1%
Haoran Wang	2014	82.2%
Our Approach		**95.5%**

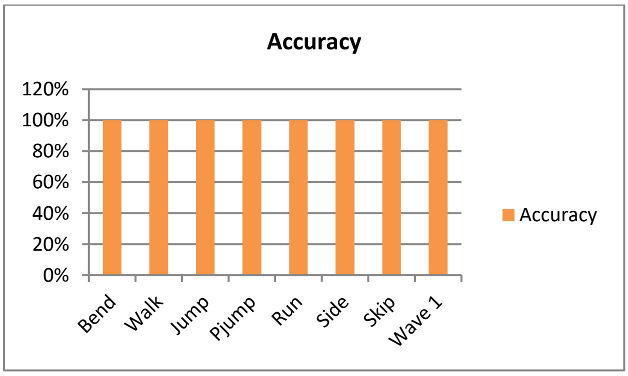
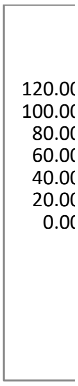


TABLE II
TABLE FOR COMPARISON ACCURACY OF DIFFERENT RESEARCHERS

Table 2: Table for comparison accuracy of different Researchers



Name of action	Accuracy
Bend	100%
Walk	100%
Jump	100%
Pjump	100%
Run	100%
Side	100%
Skip	100%
Wave 1	100%
Wave 2	100%

Fig 5.12 : Graph for accuracy table

VI. CONCLUSION

In this work, different classification techniques for human actions or activities recognition have been discussed. Each technique is better suited than the other for different types of activities in different application areas. On an average SVM performs better classification when we need a linear classification but the size of data is quite large. KNN, as discussed provides higher level of abstraction with high accuracies but time and complexity increases as compared to SVM. We also saw one more classifier 'visual words', which builds up histograms of extracted features and the classification is based on probabilistic model. Each technique has its own accuracy rate. The action classification experiments on the Weizmann dataset using KNN and SVM classifier. In the Weizmann dataset given 100% performance using static and dynamic actions, comparison is complex as little work performs the same action split. Despite this, performance of static and dynamic actions is equal, while both all and static and dynamic actions outperform approaches such as the former does not require any pre-processing which is advantageous. This research work is not extend further on the Weizmann dataset, because the accuracy of the Weizmann dataset is 100% achieved in this research work and this work was previously 98%. But our work is achieved accuracy upto 100 percentage .

FUTURE WORK

For further work, to achieve better accuracies, more than one classifiers can combined together for performing better classification and recognizing activity in the videos.

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 .