

Multi-View Feature Clustering Technique for Detection and Classification of Human Actions

Syed Thouheed Ahmed¹,Nirmala S Guptha²,AfifaSalsabil Fathima³, Ashwini S³

¹Dr. T. Thimmaiah Institute of Technology, KGF, India

²Sri Venkateshwara College of Engineering, Bengaluru, India

³Cambridge Institute of Technology, Bengaluru, India

syed@drttit.edu.in

Abstract. Recognizing the actions performed by any person is the most successful applications in pattern recognition. Detecting the action in a moving camera influences dynamic view changes, is based on spatio-temporal information at multiple temporal scales. In this paper, we are presenting a system that is dependent on actions based on multi-view information. These multi-view features are extracted from various temporal scales. The GMM and Prewitt edge filter is used for detecting background and foreground image. The Nearest Mean Classifier is used to cluster features vector's of moving object. The experiment results demonstrated using Kth dataset producing 98% of accuracy.

Keywords : Human action detection, pattern recognition, clustering, classification;

1 Introduction

Over recent years finding people actions in image sequence has attracted much attention in surveillance cameras system. It is a wide range of application in analysis of human behavior, because every object has different appearance such as object posture, shadow, different clothing or other environment conditions. It also includes various issues such as illumination, occlusion, or camera movement. The camera used in these applications, captures the object movement in static point. In case of surveillance camera, the object movement maybe an obscure position from different view point. The frames obtained from these surveillance cameras are usually with low resolution. However, there are limitations to monitor event sequence in surveillance videos. Hence, detecting the human activity offers the most vital application in pattern recognition [5].

Human action recognition has its own research field. It has gained attention towards growing security concerns. An Action is a dynamic sequence of activities executed by person for a short duration of the time. These patterns are being recorded in cameras. Usually surveillance camera is situated at distance place to convey action in signals. This recorded sequence of action is taken in uncontrolled environment with camera motion. In moving camera the object is affected by background and camera

movements. In order to extract background and foreground region GMM and prewitt edge detector is used. We use Nearest Mean Classifier (NMC) for detecting feature vector of the action in each frame. For identifying the feature vector, KTH dataset is being used. With this the system can operate at reduced computational cost [6].

In existing system the actions are obtained from static view point. The features extracted from image sequence by using optical and spatio-temporal scales. Combination of spatio-temporal features is more complex and increases the computation time. Hence, to overcome the noise challenges, many approaches are being involved. The proposed system can represent the global temporal scale information using Harris Corner detector. The multi-view feature vector for each frame is obtained using NMC to apprehend the sequence of motions at low computation cost. The features of multi-view are captured from various temporal scales using global spatio-temporal distribution features.

2 Literature Survey

V.Parameswaran et al.[1] proposed a trajectories in 2D invariance space to recognize action in general view-point. The view invariance is considered for canonical body posture. It makes use of leave one out validation technique for representing the human action in single view-point. The KTH and Weizmann dataset are being used for subject invariance. The main disadvantage is recognizing the action in 3D invariance space is difficult and increases the computation time. Steffen bickel et al.[2] presents a multi-view algorithm that optimizes agreement between the views. EM based multi-view outperform single view counter points clustering in multi-view environment. If there is no natural feature split then randomly split available features. Sanchitsingh et al.[3] the action recognition is based on silhouette image. Other image properties such as intensity of light, complexion are not being used. The disadvantage of this system is in change in object's posture and motion tracking. A MuHAVi dataset is used, which is obtained with 14 actors and 8 cameras under 17 action classes, where each and every actor perform an actions many time in various action zone. C.Thurau et al.[4] presents a technique to recognize action in video, the HOG is used to represent the basic pose of the object. An image from static camera is recorded in film. The sequence of image forms a video. This video consists of various action perform by different person under various scenario. Large amount of low dimensional local features is difficult to recognize different change in view

3 System Architecture

The proposed system the actions are being trained to substructure, based on the movements of the object. The information is trained in form of image sequence; it is necessary to segment the object from background and foreground regions. Further, the actions performed by an object are given by a boundary known as bounding boxes. The testing data of the object is provided in sequential order by point of interest to

form frames these frames are estimated by the feature vector obtained by the classifier.

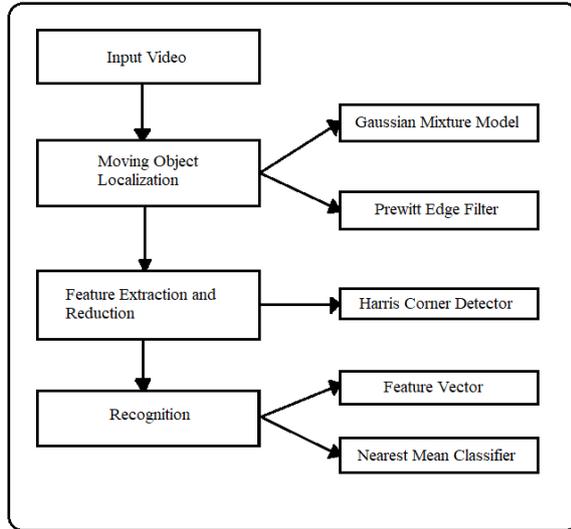


Fig.1: System Architecture

The input test videos are obtained from KTH dataset, for which the frames are extracted for subtracting the background and foreground regions. The moving object is detected and localized based on two detectors. Initially Gaussian Mixture Model is applied on the given image sequence. A low-pass filter is used in GMM model to reduce noise in the given image. Further, temporal difference is applied on those sequences to extract the silhouette image by subtracting the background image. Secondly, the prewitt edge detector is applied on the silhouette image to distinguish the edges by providing a boundary to the moving object known as bounding boxes. This boundary is very essential to locate the hands and legs of the moving object's silhouette image. The data within these bounding boxes is transmuted into set of vectors called feature vectors. For each frame the feature vector is extracted using Harris spatio-temporal corner detector, to have an equal grid for all frames a threshold value is fixed based on the view-point. The Nearest Mean Classifier calculates the mean-value for all component vectors of similar action under multiple views. On comparing the mean value with KTH dataset the action of the moving object is been identified by providing the predicted class of NMC.

3.1 Experiment Results

According to the analysis of recognition system, it is necessary to train NMC classifier; we make use of KTH datasets. This dataset is most widely used, developed at KTH Royal Institute of Technology. The video clips are developed by making use of 4 actors, performing 6 actions (walking, jogging, running, boxing, clapping and waving) under four distinctive environmental conditions(indoor, outdoor, different cloths and variations in scale) consisting of 144 video clips captured at 25FPS.

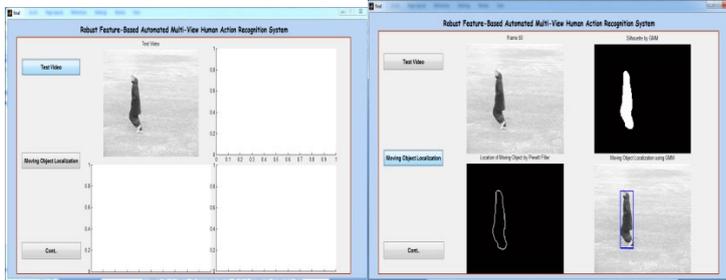
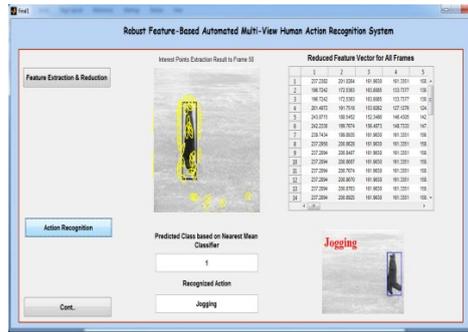


Fig. 2: (left) Input Video, (Right) Extracting Silhouette for Initial Frames and Obtaining the



Bounding Boxes

Performance Analysis		
	Proposed Method	Existing Method
Accuracy	98	94.3 %
Precision	100	98.1 %
Specificity	100	97.4 %
Recall	96	89.2 %

Fig.3: (Left) Extracting the Feature Vector and Recognizing the Human Action (Right) Comparison of Performance Analysis

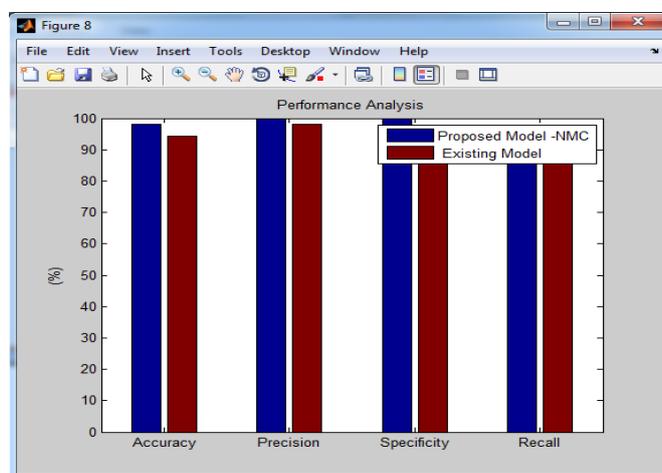


Fig.7: Graphical Rep-

resentation of Performance Analysis

The above figure demonstrates the action recognition based on the NMC classifiers. The training dataset produces an increased accuracy of 98% compared with existing system. Along with accuracy the proposed provides increased in precision time with 100% where as the existing system provides 94.3% of accuracy by combining view-invariance and subject invariance.

4 Conclusion

In this paper, the human action detection system utilizes the view-invariance feature to address the multi-view activity. The human action is recognized by extracting the background images. Each frame is regarded as a holistic feature vector value and the set of feature vector are calculated by the point of interest obtained by spatio-temporal feature values. The experiment results prove the effectiveness and robustness. The accuracy for recognition system is acquired under laboratory conditions, so it has limits. This system provides effectiveness for real-time video, thereby solving the problems caused by video quality. In future work, this recognition method will be used in surveillance field and facial expression detection.

Reference

1. V. Parameswaran and R. Chellappa, "View Invariance for Human Action Recognition," *Int'l J. Computer Vision*, vol. 66, no. 1, pp. 83101, 2006.
2. Steffen bickel and tobiasscheffer department of computer science Humboldt university "multi-view clustering" proceedings of fourth IEEE international conference on data mining.
3. Sanchit Singh, Sergio A Velastin , DIRC, Kingston University Kingston upon Thames, UK HosseinRagheb ISBE, University of Manchester Manchester, UK "MuHAVi: A Multicamera Human Action Video Dataset for the Evaluation of Action Recognition Meth-

ods” Seventh IEEE International Conference on Advanced Video and Signal Based Surveillance.

4. C.ThurauandV.Hlavác,“Pose primitivebasedhumanactionrecognition in videos or still images,” in Proc. IEEE Conf. Comput. Vis. Pattern Recog., Anchorage, AK, 2008, pp. 1–8.
5. Kumar, S.S., Ahmed, S.T., Vigneshwaran, P. et al. “Two phase cluster validation approach towards measuring cluster quality in unstructured and structured numerical datasets”. J Ambient Intell Human Comput (2020). <https://doi.org/10.1007/s12652-020-02487-w>
6. Ahmed, S.T., Sankar, S. & Sandhya, M. “Multi-objective optimal medical data informatics standardization and processing technique for telemedicine via machine learning approach”. J Ambient Intell Human Comput (2020). <https://doi.org/10.1007/s12652-020-02016-9>
7. Thouheed Ahmed S., Sandhya M. (2019) “Real-Time Biomedical Recursive Images Detection Algorithm for Indian Telemedicine Environment”. In: Mallick P., Balas V., Bhoi A., Zobaa A. (eds) Cognitive Informatics and Soft Computing. Advances in Intelligent Systems and Computing, vol 768. Springer, Singapore. https://doi.org/10.1007/978-981-13-0617-4_68
8. S. T. Ahmed, "A study on multi objective optimal clustering techniques for medical datasets," 2017 International Conference on Intelligent Computing and Control Systems (ICICCS), Madurai, 2017, pp. 174-177, doi: 10.1109/ICCONS.2017.8250704.