International Research Journal of Engineering and Technology (IRJET) e-ISSN: 2395-0056
Volume: 08 Issue: 09 | Sep 2021 www.irjet.net p-ISSN: 2395-0072

# Object's Action Recognition System based on Multi-View Feature

**Mithun NS[1], Harini S[2], Hemashree K[3], Preethi M[4], Ashwini S[5]**

[1-4]*Diploma, Department of CSE, Bharath Polytechnic, Karnataka, India*
[5] *Lecture, Department of CSE, Bharath Polytechnic, Karnataka, India*

---

***Abstract -*** Image processing is the digital image that is used in computer algorithms, and it consists of the manipulation of images by using a digital computer. In Existing system, the object movement is under static viewpoint. The action has effect on the viewpoint. Single classifier it is used to train the system and each classifier manages every viewpoint. The existing system is more sensitive to noise. The proposed system presents the sequence of the activities, and we have all human action. The proposed system approaches the delineated to confess the action under multi-viewpoint and in this system NMC Model is used to the multi view features are extracted from the scales and in this System, videos are the acquiring from database and frames of extraction is used for performing the further process. The motivation includes the additional layer of variability the activity of caught under limited and inferred situation. The system does not to be retrained when scenarios are changed which means the trained database can be applied in a wide variety of environment such as background changes and the database can be applied on any platform.

***Key Words***: Background Detection, GMM and Prewitt Edge Detector, Harris Spatio-Temporal Corner Detector, Nearest Mean Classifier (NMC), Kth Dataset

## 1. INTRODUCTION

There is a rapid development in the society of many smart phones and the application via internet, Even the Digital multimedia are favored nowadays because it is relating the ability to interpret or to become aware of something through the senses. We can see the advanced technology development and it performs a vital role in the research.

Human action recognition is very much inspected in many applications, and it is effectual action recognition it indicates the action and details are favorable outcome and pictures. The correct pursuit information is taken from different media files like pictures videos. Many infinite movements have been pointed and extracted from the video, every point has been pointed and every section includes the calculative histogram vectors. At those highlights the recordings are forthright; the activities have an acceptance of strategies to execute the proper result. However, the noises don't interact with data that might be added, and the close-up highlights will be extracted, hence the present techniques or ideas are generally not capable or influential and they couldn't be applied when the video is mix up in a confused state or untidy way, barrier, jiggle, etc. So, to enrich the acceptance heed and imperative ideas, it will give an exposition of human activities and understanding it.

These ideas may require some procedures to take a step to record and to do execution, mainly for the correct videos. In many ways, human action is tendency that is passed by the pictures. These activities adapt the spaces among the various kinds of features. Typically, many highlights have been amplified along with the attributes. For the present, the adapted thing is expanded and converted into many diverse spaces. To investigate the close- up intricate structures and videos are used to enable information in the video domain. The results of the test cases will represent the calculative way and that isn't just effectual it also works better in execution. Multi view of human action recognition system uses Mat lab and helps in image processing ideas and classification of algorithms. These reports are taken from the base papers where it is completed by reasonable information and a set of collection and it ignores the low-quality pictures (rarities, noises) and more troubles like (obstruction, shade, in motion of pictures, varied garments) to know their results are perfect. The frame will be working nonstop which is easy, rapid and identifies a fall. In addition, the settings are the foundation, camera position, camera movement, cooperation among the people as an individual.

The strategies are investigated to tend to be in many different vision issues. The calculation is presented, and every track is clearly noted by the view. At that time, the combination joins the investigation from the posture classifier, the existing model supports the limitation of the model and the goal of the projection which matches camera sees. Considering, the difficulty to get named for the video set data, there are many applications to create model adaptable chart by contiguous network configure. The domain adapter method uses the semi-directed attributes to increase the information in given space. Although, they are specifically used in many applications of existing techniques. And this is not described because the video contains some disadvantages like camera shakes and blurred background. The results display that the actions are present in the datasets and genuine world datasets. And it demonstrates the techniques outperform. The methodology of proposed system has been explained

in this section and the outcome of results are discussed and illustrated here.

## 2.LITERATURE SURVEY

V.Parameswaran et al. [1] proposed a trajectory in 2D invariance space to recognize action in general viewpoint. 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 viewpoint. 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. Sanchit singh et al. [3] the action recognition is based on silhouette image. Other image properties such as intensity of light, complexion is 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 action 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. F.Murtaza et al.[5] proposes a HOG for detecting multiple humans in the given video sequences, accurate for both visually and empirically produces sequence. It provides accurate identification for moving regions. The occlusions are not being handled due to multiple object motion. Identifying the changes in multiple object movement is difficult due to the variations in brightness. Kel et al. [6] presents a Volumetric Feature for decomposing the actions into optical flow. A VOL boost classifier is utilized for the estimating of optical flow, to build action templates for the optical flow. Decomposing the optical flow into horizontal and vertical component consumes large amount of computation time. Calculating the volumetric feature for multiple objects and decomposing features for such objects is complex. P.Scovanner et al.[7] provides a compact representation of primitive action. A SVM split detector is used for representing actions using spatio-temporal features for producing Histograms. The view-point estimation is used for view-invariant action recognition and also it can deal with complex feature by making use of code book approach for object recognition. There is a complexity in using 2D spatio-temporal feature removed output is done, then Histogram output is obtained, then Thresholding technique is applied.

## 3. SYSTEM ARCHITECTURE

In this proposed work, the video is given as an input, and it is converted into number of frames(fig.1), so these frames are calculated by GMM model. Considering computational productivity, it will separate are extracts Key edges. Now the video features are extracted in the form of images, and it is taken before to classify using classification techniques. The color histogram has been determined and subtracted in frames. The subtracted values are more significant than the set then the frames is shot boundary and the shot is known as key frame when it is taken. Filtering techniques is a preprocess to remove the noises in the frame, by this extraction method it will reduce the storage area in the dataset from this the output can be sorted by comparing test video frames in which value extracted with approximate dataset values using the above procedure or techniques. NMC classifier proves to be the exquisite in the classification of all human action recognition.

KTH dataset(fig.2) is one among the popular datasets in human action recognition. This dataset consists of six actions, viz., walking, jogging, running, boxing, hand-waving, and handclapping which were carried out by 25 persons and the videos were recorded in four different scenarios (outdoor, variations in scale, variations in cloths, and indoor). The spatial dimension of each frame is 160 × 120 pixels and the rate of frames per second (fps) is 25. This dataset has 600 videos. All the videos were captured from a distance from the performer. As a result, the area covered by the person is less than 10% of the whole frame. To account for performance nuance, each action is performed by 25 different individuals, and the setting is systematically altered for each action per actor.

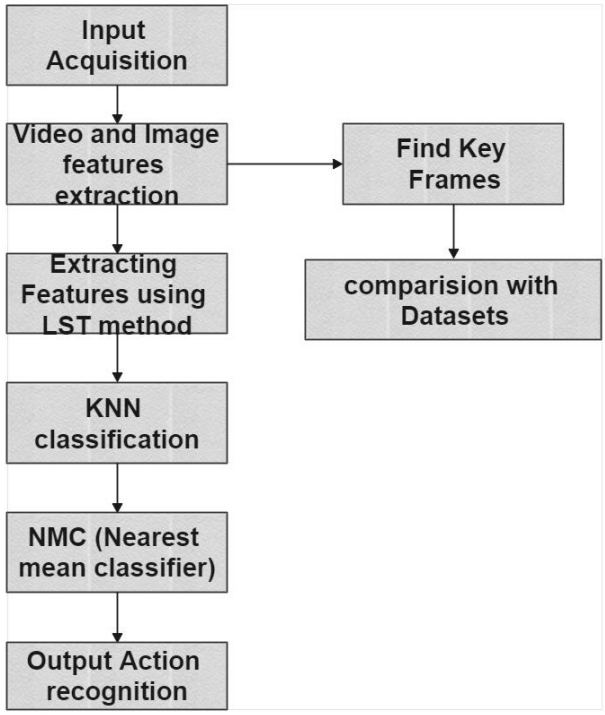


Fig.1: System Architecture

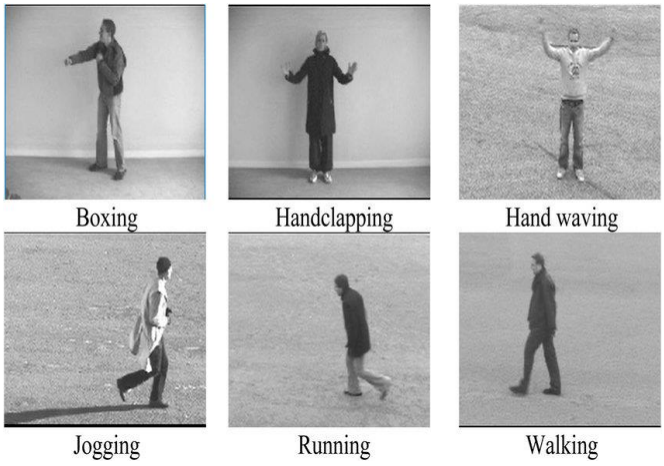


Fig. 2: Kth Dataset

## 4. MATHEMATICAL MODEL

**Video Acquisition:** Collect the frames (F) from the sequence of Image (I). Considering incoming frames F where F= {f1, f2----fn} where n=50 Such that

Fi ∈ K / K ≠ Null------(1) Where K is feature Set

***Moving Object Localization:***

Let M ∉ Null and M ∈ Fi

$$M = \sum_{i=0}^{n} \frac{\text{Gi}}{\text{Pi}} \text{ ----------(2)}$$

Consider the frames (F) while subtracting the background from image

$$\mathbf{k}= R\left[{}_{0}^{m}(k_j)_i\right] \text{ ----------(3)}$$

Where k is background image

R is range of feature vector.

K is internal feature set.

Consider the Pi for subtracting Foreground Image

$$P_i = (\sum_{i=0}^{n}. \sum_{j=0}^{m} R[{}_{0}^{m}(k_j)_i) \text{ ----------(4)}$$

**Feature Extraction**

$$E = \sum_{i=0}^{n} \frac{d(M_i)}{dF}$$

Where i is the number of frames from 1 to 50 and E is the Average of $\frac{\text{d}(M_i)}{\text{dF}}$ where E is $\frac{\partial(Mi)}{\partial F}$

$\left[\frac{\partial(M_i)}{\partial F}\right]_0^n$ where E ⊆ D where D is Kth dataset

on comparison of E and Kth Dataset,

if E ⊆ D is true then A= $\frac{\text{E} \subseteq \text{D}}{600} * 100$

Where A is compared value of E and D

## 5. EXPECTED RESULTS

Once the user opens the MATLAB and run the code, it pops up a window which interrogate the user to select the input. User selects the input video and the selected video runs, is shown in figure 8.2

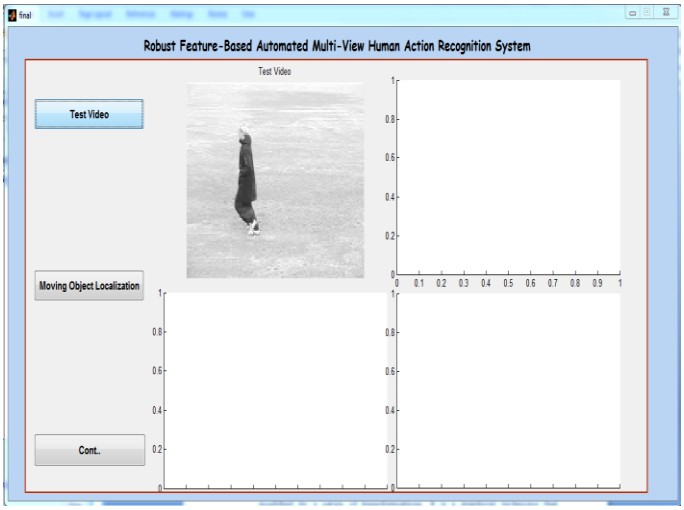


Fig. 3: Input Video

The algorithm runs by calling GMM to extract the silhouette of the moving object by subtracting the background. Further, the Prewitt edge filter subtracts the foreground image, and the bounding box is obtained against the moving object(fig.3). Silhouette is obtained at final frames to order to get the complete silhouette and bounding boxes across the object. In this step the frames have reached the threshold state to check whether the object has its complete bounding box and silhouette of the object. Feature vector for all frames as the object is moving.

Once the frames have reached the threshold value the feature vector for all the frames is obtain as the histogram range. The action is confessed using Nearest Mean Classifier (NMC), is shown in fig.4

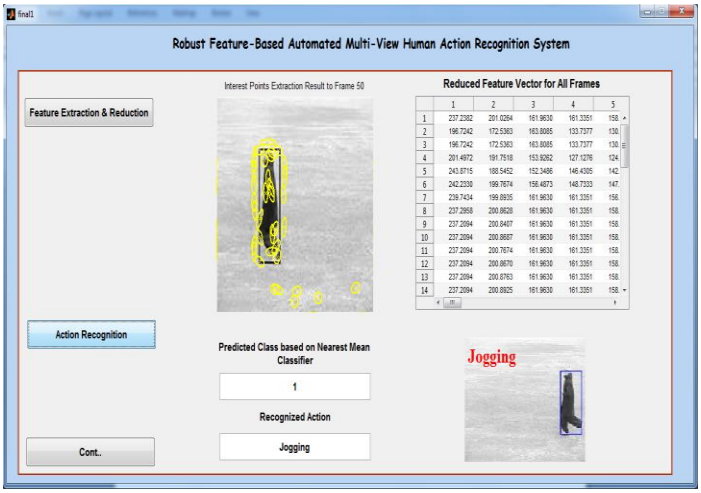


Fig. 4: Recognizing the Human Action

The performance is calculated against the proposed and existing system. The existing system uses cross data testing on Kth, WEIZMANN and MuHAVi datasets whereas the proposed system uses Kth data set compared with Nearest Mean Classifier (NMC), is shown in figure 5 and 6

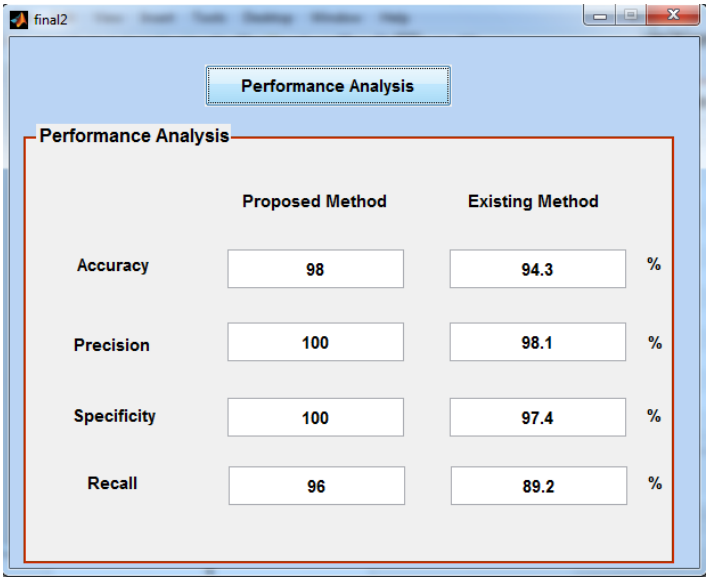


Fig. 5: Comparison of Performance Analysis

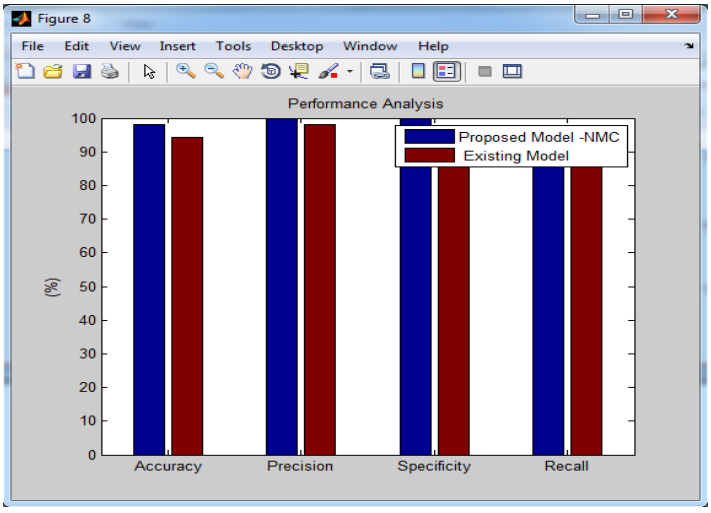


Fig. 6: Graphical Representation of Performance Analysis

## 6. CONCLUSION

In this survey it presents advanced techniques of automatic human action recognition which includes some basics of image processing, acquisition, enhancement in the quality, value or extent, image compression and restoration it gives an effective improvement and some gestures to recognize so these are discussed in the paper. It is tested by some action recognition in proposed system and the execution is resulted. It observes the class improvement by classifying together with datasets. The survey demonstrates that actions of the human by recordings or images. It makes effortless when it comes to research of fault detection.

The reputation accuracy turned into obtained beneathneath laboratory setting, so it had a few limits. The future of studies is evaluating the consequences with present system. This class approach could be utilized in intelligence surveillance area the usage of dynamic background, adopting a stepped forward version primarily based totally on localized components of the image.

## REFERENCES

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 tobias scheffer 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 Hossein Ragheb ISBE, University of Manchester Manchester, UK "MuHAVi: A Multicamera Human Action Video Dataset for the Evaluation of Action Recognition Methods" Seventh IEEE International Conference on Advanced Video and Signal Based Surveillance.
4. C.Thurau and V.Hlavác, "Pose primitive based human action recognition in videos or still images," in Proc. IEEE Conf. Comput. Vis. Pattern Recog., Anchorage, AK, 2008, pp. 1–8.
5. F Murtaza, MH Yousaf, SA Velastin, "Multi-view human action recognition using 2D motion templates based on MHIs and their HOG description". IET Comput. Vis. **10**(7), 758–767 (2016)
6. Yan Kel, Rahul Sukthankar, and Martial Hebert. "Efficient visual event detection using volumetric features." In ICCV, pages 166 – 173, 2005.
7. P. Scovanner, S. Ali, and M. Shah, "A 3-Dimensional Sift Descriptor and Its Application to Action Recognition," Proc. Conf. Multimedia, pp. 357-360, 2007.
8. J Jiang et al., "Human action recognition via compressive-sensing-based dimensionality reduction." Optic-International Journal for Light and Electron Optics **126**(9), 882–887 (2015)
9. Kuang-Pen Chou, Mukesh Prasad, Di Wu, Nabin Sharma, (Senior Member, Ieee), Dong-Lin Li, Yu-Feng Lin, Michael Blumenstein, Wen-Chieh Lin, And Chin-Teng Lin," Robust Feature-Based Automated Multi-View Human Action Recognition System." Special Section on Visual Surveillance and Biometrics: Practices, Challenges, and Possibilities, Volume 6, pp 15283- 15296, 2018.