Available online at www.sciencedirect.com

ScienceDirect

Procedia Computer Science 167 (2020) 870–877

Procedia Computer Science

www.elsevier.com/locate/procedia

International Conference on Computational Intelligence and Data Science (ICCIDS 2019)

Robust Human Action Recognition System via Image Processing

U. Anitha[a]*, R.Narmadha[b], D. Raja Sumanth[c], D. Naveen Kumar[c]

[a]*Assistant Professor, Department of ECE, Sathyabama Institute of Science and Technology, Chennai, India,*
[a]**Email Id:anithaumanath@gmail.com*
[b]*Associate Professor, Department of ECE, Sathyabama Institute of Science and Technology, Chennai, India*
[c]*U.G. Students, Department of ECE, Sathyabama Institute of Science and Technology, Chennai, India*

Abstract

Human actions detection is very much investigated in utilization of artificial intelligence and computer vision. Numerous effective action recognition strategies have demonstrated and the action information are successfully gained from motion videos and still pictures. In order to get the equivalent actions, the proper activity information gained from various kinds of media like videos or pictures might be connected. The majority of the existing video activity action identification strategies experience the ill effects of inadequate marked recordings. In that cases, over-fitting should be a potential issue and the execution of activity acknowledgment is controlled. In this paper, image processing techniques are used in order to recognize the different hand poster of the human body, also the over-fitting can be eased and the execution of activity acknowledgment is improved. Initially, the human action video including hand waving, walking, jogging, clapping, boxing is converted into image of 2D frames and then it is preprocessed followed by feature extraction using LST and classification by KNN classifier has been done individually. The kernel principal component analysis (KPCA) technique is used in the proposed system for finding the image features and joined features.The extracted features from the frames are compared with trained quantized dataset in order to identify the actions. The advantage of quantized dataset is that it occupies very less space. Thus, the result shows which action is present in the examined data. Trials on open benchmark data sets and genuine world data sets demonstrate that our technique outflanks a few other cutting edge activity acknowledgment strategies.

Keywords:Human action recognition; Image processing; Bi-linear interpolation;Laplace smoothening transform (LST); KNN Classifier

10.1016/j.procs.2020.03.426

1. Introduction

The fast development of internet applications and smart phone, an action acknowledgment in private videos delivered by clients are turn into a vital research theme because of their large applications, such as programmed video tracking, image annotation and so on [1]. Accordingly, these videos contain extensive in to a class varieties inside the equivalent semantic classification. It is currently a testing assignment to perceive human activities in such videos. Many activity acknowledgment techniques pursued the customary system. Initially, countless movement highlights are extricated from videos. At that point, every single neighborhood include are quantized in to a histogram vector utilizing back of-words (bow) portrayal. Later, the vector-based-classifiers, e.g., bolster vector machine are utilized to perform acknowledgment in the testing and recording. At a point when the recordings are straightforward, these activity acknowledgment strategies have accomplished promising outcomes. Nonetheless, noises and the uncorrelated data might get added to the bow amid the quantization and extraction of the nearby highlights [9]. In this way, these techniques are typically not powerful and couldn't be used much when the video having significant camera shaking, impediment, jumbled foundation, etc. So as to improve the acknowledgment precision, important parts of activities, e.g., related articles, human appearance, act, etc, ought to be used to form a clearer semantic understanding of human activities. Late endeavors have exhibited the viability of utilizing related items or human postures. These techniques may require a preparation procedure with extensive measure of recordings to get great execution, particularly for true videos. In most cases, human activity inclination can likewise be passed on by still pictures [6, 7]. In proposed an adjustment technique for video action recognition. Not quite the same as the current adjustment methods based on a similar component, our strategy can able to adapt knowledge among spaces that are in various feature spaces. Distinctive highlights can give enhanced performance and thanks to the corresponding attributes.

Meanwhile, the adaptability expanded and the adjustment can be conducted between diverse spaces. In request to investigate the nearby complicated structures along with the preparing video information successfully use the unlabeled information in video domain, the adjustment procedure in a semi supervised learning system can be done [10]. Test results show that the calculation isn't just effective but also has better adjustment execution, particularly when just few named preparing tests are given.

The past and ebb and flow inquire reports about robust feature-based automated multi view human action recognition system with the help of the image processing techniques and classification algorithms using Matlab have been contemplated [11-14]. Every one of these reports is taken as a base for this paper. Caroline Rougier, et.al, were finished their work with a reasonable informational collection, and in dislike of the low-quality pictures (high pressure ancient rarities, noise) furthermore, division troubles (impediments, shadows, moving objects, diverse garments, etc), and acknowledgment results are good. The framework can keep running continuously at 5 outlines which are quick and adequate to identify a fall. At last, looked at with other 2-d includes, the shape disfigurement highlights are fundamentally better devices than distinguish falls when growing such frameworks. This necessity is happy with our framework as it is totally robotized, and no one can approach the pictures aside from if there should arise an occurrence of crisis [8]. The framework will be actuated to send an alert flag toward an outside asset (e.g., by means of a mobile phone or internet) if and just if an irregular occasion is recognized (e.g., falling). In addition, this is a strategy that does not require the individual to wear any gadget. Ronald Poppe, et.al, were discussed about vision-based human activity acknowledgment in this review yet a multi-modular methodology could improve acknowledgment in a few areas, for instance in motion picture investigation. Additionally, setting such as foundation, camera movement, association among people and individual character gives enlightening signs. Given the present best in class and spurred by the expansive scope of utilizations that can profit by vigorous human activity acknowledgment, it is normal that a large number of these difficulties will be tended to sooner rather than later. This would be a major advance towards the satisfaction of the longstanding guarantee to accomplish vigorous programmed acknowledgment and translation of human activity [2].

In next audit, they condensed the principle strategies that were investigated for tending to different vision issues. The secured themes included item following an acknowledgment, human action investigation, hand motion

examination, and indoor 3-d mapping. They additionally proposed a few specialized and scholarly challenges that should be examined later on [3]. The proposed calculation distinguish, track, and concentrate includes freely in each view. At that point, a combination unit blends the stance investigation to give a standing/stretched posture classifier that is productive in unspecified perspectives and falling headings. From the posture probability estimation, the induction is performed with respect to every one of the cameras together, and is overseen by utilizing a Layered Hidden Markov Model (LHMM). This affiliation manages unexpected changes furthermore, is strong to low-level mistakes [4]. The study uncovers essential advancement made in the most recent ten years in little vocabulary, single-individual, full-body activity acknowledgment. Imperative issues that must at present be tended to in future work are versatility of activity acknowledgment frameworks as for vocabulary measure; acknowledgment within the sight of obscure activities; scenes containing various people; and connections between products people [5].

To start with, the existing model supports a great 3- d limitation of the model with the end goal that its projection matches the inside and out camera sees. This is great news for the possibility of any multi-see 3-d model based following technique. Since it is much difficult to get a well-labeled video data, it needs semi-supervised procedure to employ unlabeled videos. To use the complex structure of mutually marked and unlabeled preparing information, there are many existing methods in the present world. One of method used is a semi-supervised discriminant analysis (SDA) system which has been utilized by bringing the geometrical regularize into the ideal function of Linear Discriminant Analysis (LDA). Additionally the created model of adaptable chart by coupling stay based mark expectation and contiguousness network configuration has been used in different applications. These robust semi-directed strategies, the laplacian lattice can be learned by utilizing neighborhood relapse and worldwide arrangement and these are likewise being used in different applications.

For the purpose of domain adaptation methods, additionally utilize this semi-directed element increase with the unlabeled information in the objective space. Although these are used in various applications there are few drawbacks present in the existing techniques. It is not strong and might not be described well when the video contain considerable camera shakes and cluttered background. Need large dataset for training. The proposed system is separated into two parts: online training and offline testing. The feature extraction is the first step in offline training. The feature vectors of each image in sequence are described and are quantized to reduce their dimension. Finally, they are stored in the database. The first two steps of online training is similar to offline training. The dimension of the feature vector is decreased by using the histogram technique. The result shows that the action is present in the test data. Trials on open benchmark data sets and genuine world data sets demonstrate that our technique outflanks a few other cutting edge activity acknowledgment strategies. The proposed system is explained in section 2 and the methodology, results and discussions are illustrated in section 3 and 4 respectively.

2. Proposed System

In the proposed work, the image feature from the pictures and key edges of videos were extracted. Considering computational productivity, the proposed system will separate key edges by a shot boundary detection algorithm. First the video is given as input and then the features are extracted in the form of images and then combined with video feature and preceded to classifier and by using classification techniques the output is generated as shown in figure 1.

To start with, the colour histogram of each 5 frames is determined. Second, the histogram is subtracted with that of the earlier frame. Third, when the subtracted value is bigger than the empirically set threshold then the frame will be set as a key frame shot boundary. The frame in the center of the shot is considered as a key frame only when we get the shot. This method is called shot boundary detection. Meanwhile, the video (movement) is separated from the video domain and joined with the image feature. The picture element is a subset of the combined element. The Kernel Principal Component Analysis (KPCA) technique is used in the proposed system for finding the image features and joined features.

The KPCA strategy says the primary information of the mapped hilbert spaces. In this way, the preparation procedure is progressively proficient, which makes the Independent Vector Analysis (IVA) increasingly reasonable for true applications. The common features can be obtained by mapping the image feature into a hilbert space. So as to get the heterogeneous features-ab, the joined features are mapped into another hilbert space. The information can be adjusted dependent on those shared space with the common features, after it is used to upgrade the classifier-a. So as to make utilization of unlabeled videos, a semi supervised classifier-ab is prepared dependent on the heterogeneous features in video domain. By combining the two classifiers we can get joint optimization framework. The last acknowledgment after effects of testing recordings are improved by combining the consequences of previously mentioned two classifiers. It avoids over fitting and gives good performance even during a few labeled training videos are available.

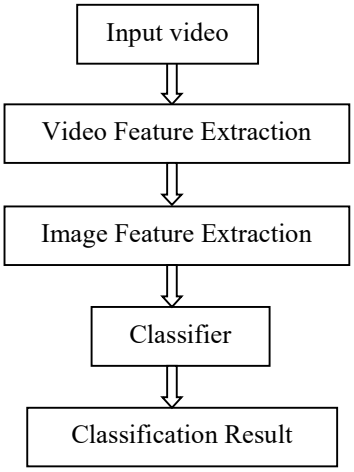

Fig.1 flow chart of the proposed work

3. Methodology

The video is given as input to the video reader and converted them into number of frames. The frames are then calculated in the form of histograms and smoothened them by using Laplace smoothening transform (lst).The colour histogram of each five frames is determined and subtracted with that of the previous frame. If the subtracted value is bigger than the empirical set threshold value then the frame will be shot boundary. The frame in the center of the shot is considered as a key frame only when the shot is taken. The noises in the frames are removed by using preprocessing (filtering) technique. The frames will be in the form of 3D image and it must be converted into 2D image for the better result of the given input video by using the bi-linear interpolation process. By using the feature extraction method, the frames from both the image and video features can be extracted clearly by using Laplace smoothening transform. The quantized trained dataset is used in this system will reduce the storage area. The output can be classified by comparing the extracted value of the input test video frames with the quantized dataset values using the above process.

The KNN classifier proves to be best in the classification of human action such as hand waving, walking, jogging, clapping and boxing.

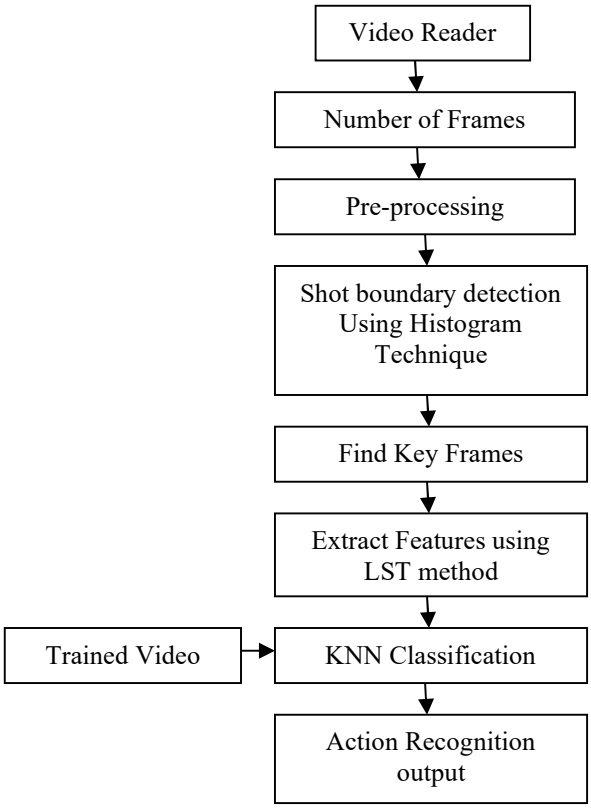

Fig.2 process of the hand recognition system

Histogram

It represents the different frequencies of the image pixels which is used for the understanding of its features.

Bi-linear interpolation

This is a resampling method which uses the distance weighted average of the four nearest neighbor pixel values to estimate a new pixel value. It is used here for the resizing of the image.

Laplace smoothening transforms (LST)

The LST transform is proposed for extracting low frequency characteristics of the image based on laplacian function.

K Nearest Neighbor classifier (KNN)

The KNN classifier is more widely used for solving classification problem. The algorithm works based on Euclidien distance formula. The k value is chosen as 5 normally for all the applications.

4. Results and discussions

The proposed method is developed by using Matlab. In the proposed method only the action of the single person can be generated. In a video a sample of five frames were taken and a key frame is identified by shot boundary method. The process is repeated continuously and the key frames features are extracted and classified by the proposed system as shown in the following figures.

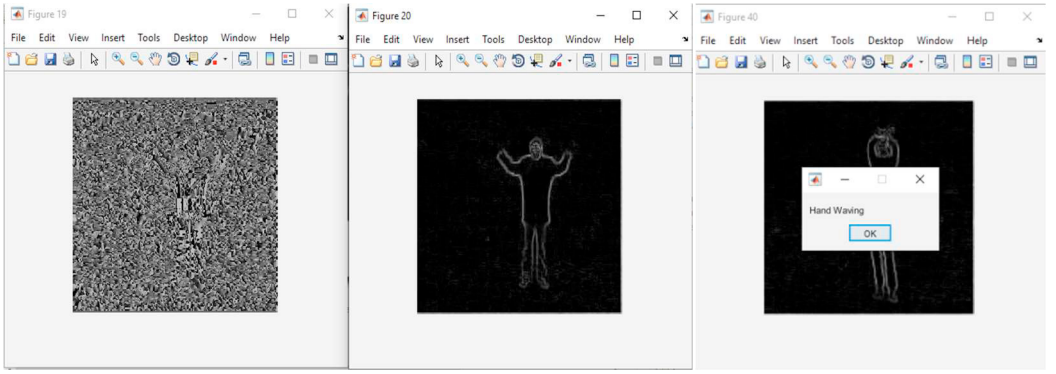

Fig.3 test output for hand waving

The first image in figure.3 represents the noise free output and proceeds to feature extraction in the second image then the classification of the test output of a hand waving action is viewed in the last image.

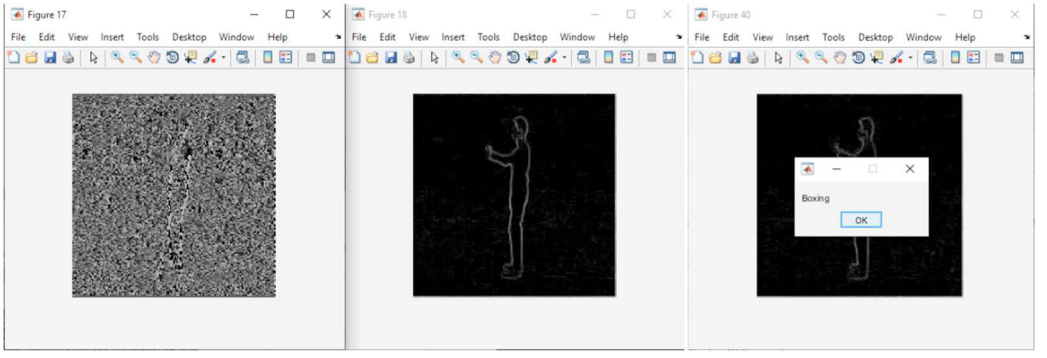

Fig.4 test output for boxing

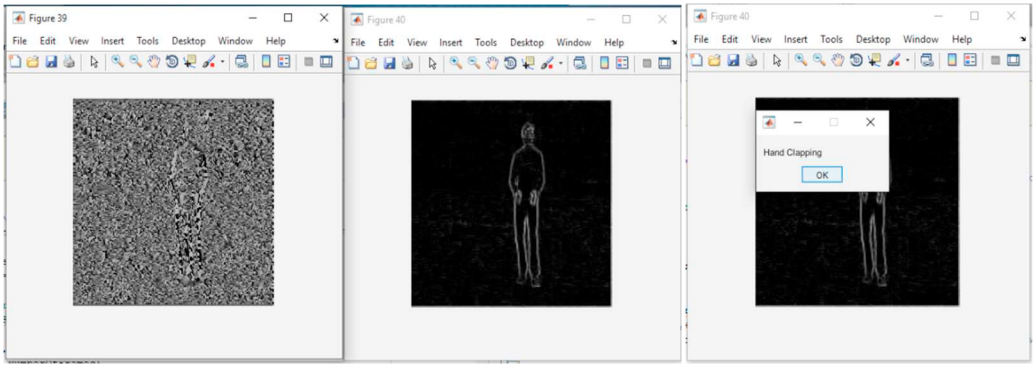

Fig.5 test output for hand clapping

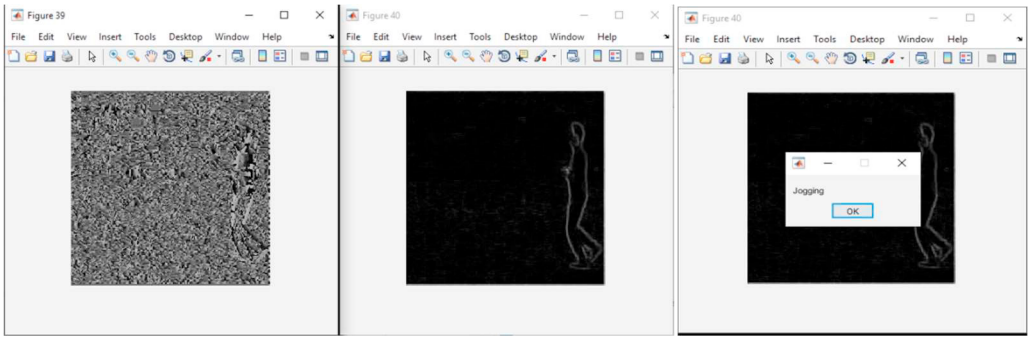

Fig.6 test output for jogging

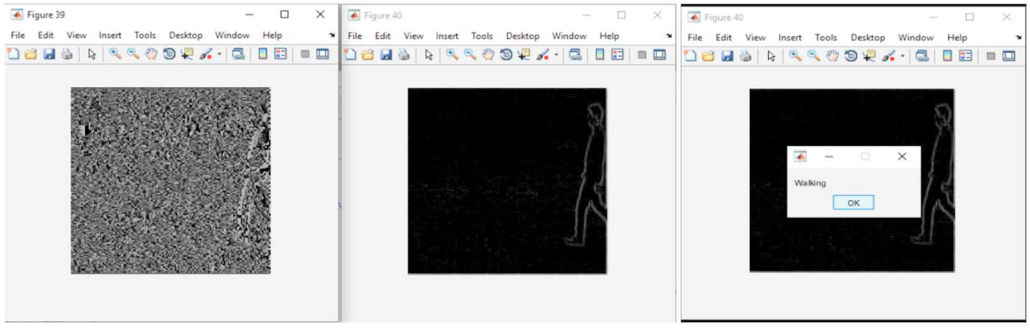

Fig.7 test output for walking

Similar to figure3, all other human actions were executed and its results are given in figure. 4 (boxing), figure.5 (hand clapping), figure.6 (jogging), figure7 (walking). The proposed method gives more efficient result for the human action recognition problem with less storage space for the trained data set.

5. Conclusion

A video action recognition system is proposed for five different hand posters and its results are executed. Test results shows that the projected system has improved execution of video action recognition, compared with old techniques. In the proposed method only the action of the single person can be generated. In future work, the actions of multiple persons in the given input video can be generated. In the proposed system test results shows the information gained from images can impact the recognition exactness of videos.

References

[1] Caroline Rougier, et.al, "Robust Video Surveillance for Fall Detection Based on Human Shape Deformation", IEEE Transactions on Circuits and Systems for Video Technology, Vol. 21, No. 5, May 2011. pp. 611-622.

[2] Ronald Poppe, "A survey on vision-based human action recognition", The Netherlands Image and Vision Computing, vol. 28 (2010), Pp. 976–990.

[3] Jungong Han, et.al, "Enhanced Computer Vision with Microsoft Kinect Sensor: A Review", IEEE Transactions on Cybernetics, Vol. 43, No. 5, October 2013. Pp. 1318 – 1334.

[4] Nicolas Thome, et.al, "A Real-Time, Multiview Fall Detection System: A LHMM-Based Approach", IEEE Transactions on Circuits and Systems for Video Technology, Vol. 18, No. 11, November 2008. Pp. 1522-1532.

[5] Daniel Weinland, et.al, "A survey of vision-based methods for action representation, segmentation and recognition", Computer Vision and Image Understanding, vol.115 (2011), Pp. 224–241.

[6] D.M. Gavrila and L.S. Davis "3D model-based tracking of humans in action: a multi-view approach", IEEE. Pp. 73-80.

[7] B. Ma, L. Huang, J. Shen, and L. Shao, "Discriminative tracking using tensor pooling," IEEE Trans. Cybern., to be published, doi: 10.1109/TCYB.2015.2477879.

[8] L. Liu, L. Shao, X. Li, and K. Lu, "Learning spatio-temporal representations for action recognition: A genetic programming approach," IEEE Trans. Cybern., vol. 46, no. 1, Jan. 2016, Pp. 158–170.

[9] A. Khan, D. Windridge, and J. Kittler, "Multilevel Chinese takeaway process and label-based processes for rule induction in the context of automated sports video annotation," IEEE Trans. Cybern., vol. 44, no. 10, Oct. 2014, Pp. 1910–1923.

[10] H. Wang, M. M. Ullah, A. Klaser, I. Laptev, and C. Schmid, "Evaluation of local spatio-temporal features for action recognition," in Proc. Brit. Mach. Vis. Conf., London, U.K., 2009, Pp. 124.1–124.11.

[11] L. Shao, X. Zhen, D. Tao, and X. Li, "Spatio-temporal Laplacian pyramid coding for action recognition," IEEE Trans. Cybern., vol. 44, no. 6, Jun. 2014, Pp. 817–827,.

[12] M.-Y. Chen and A. Hauptmann, "MoSIFT: Recognizing human actions in surveillance videos," School Comput. Sci., Carnegie Mellon Univ., Pittsburgh, PA, USA, Tech. Rep. CMU-CS-09-161, 2009.

[13] M. Yu, L. Liu, and L. Shao, "Structure-preserving binary representations for RGB-D action recognition," IEEE Trans. Pattern Anal. Mach. Intell., to be published, doi: 10.1109/TPAMI.2015.2491925.

[14] L. Shao, L. Liu, and M. Yu, "Kernelizedmultiview projection for robust action recognition," Int. J. Comput. Vis., 2015, doi: 10.1007/s11263-015-0861-6.