International Journal of Scientific Research in Engineering and Management (IJSREM)
Volume: 07 Issue: 07 | July - 2023 SJIF Rating: 8.176 ISSN: 2582-3930

"HUMAN ACTION RECOGNITION SYSTEM USING MATLAB"

Mr. Dhananjay A. Deshpande, Prof. Snehil G. Jaiswal

MTech Student, Department of Electronics and Telecommunication, G. H. RAISONI University, Amravati, Maharashtra, India
Assistant Professor, Department of Electronics and Telecommunication, G. H. RAISONI University, Amravati, Maharashtra, 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 pre-processed 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.

Key Words: Action Recognition System, Convolutional Neural Network (CNN), Human Activity Recognition (HAR), Deep Learning (DL), Machine Learning (ML), 3D CNN.

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 Accordingly, these videos contain extensive in to a class variety 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. 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.

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. 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.

In next audit, they condensed the principal 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. The proposed calculation distinguishes, 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. 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.

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.

2. LITERATURE REVIEW

For many years human action recognition has been studied well. Most of the action recognition methods require to manually annotate the relevant portion of the action of interest in the video. In recent years it has been studied that the relevant portion of action of interest can be found out automatically and recognize the action. We can review the action recognition methods.

A. Action Recognition

For representing video, feature trajectories have shown efficiency. But the quality and quantity of these trajectories were not sufficient. As the use of dense sampling came popular for image classification Wang et al. [1] proposed to use dense trajectories for representing videos. Dense points from each frame are sampled and traced them based on displacement information. For improving the performance Wang et al. [2] takes into account the camera motion.

The camera motion is estimated by matching feature points between the frames by using SURF descriptors and dense optical flow. Another approach [3] aimed at modelling the motion relationship. The approach operates on top of visual codewords derived from local patch trajectories, and therefore does not require accurate foreground-background separation. Dorr et al. [4] proposed another method for finding the informative regions. They used saliency mapping algorithms. As a new method this paper proposes using a joint learning framework for learning spatial and temporal extents of action of interest. sentence, it is spelled out. When occurring in the middle of a sentence, these words are abbreviated Sec., Ref., Eq., and Fig.

At the first occurrence of an acronym, spell it out followed by the acronym in parentheses, e.g., charge-coupled diode (CCD).

B. Action Detection

Recognition was performed using the Mahala Nobis distance between the moment description of the input and each of the known actions. Recent popular methods which employ machine learning techniques such as SVMs and AdaBoost, provide one possibility for incorporating the information contained in a set of training examples.[4] introduces the Action MACH filter, a template-based method for action recognition which is capable of capturing intraclass variability by synthesizing a single Action. Another method is proposed in [5], multiple-instance learning framework, named SMILE-SVM (Simulated annealing Multiple Instance Learning Support Vector Machines), is presented for learning human action detector based on imprecise action location. Wang et al. [6] used a figure-centric visual word representation. In that localization is treated as latent variable so as to recognize the action.

A spatio-temporal model is learned. During the training [7] model parameters is estimated and the relevant portion is identified.[8] proposed an independent motion evidence feature for distinguishing human actions from background motion. Most of the methods require that the relevant portion of the video has to be annotated with bounding boxes. Human intervention was tedious. So, to overcome the bounding box Brendel et al. [9] divides the video into a number of subgroups and then a model was generated that identify the relevant subgroup. This paper introduces a method that learns both spatial and temporal extents for detection improvement. Dense trajectory is used here as local features to represent the human action.

C. Same Domain Related Work Analysis

Several recent depth-based approaches have been reported to improve human action recognition accuracy. An action graph based on a sampled 3D representation from a depth map to model the human motion is proposed in. Several 4D descriptors have been used to represent the human action. In a histogram of oriented 4D normal (HON4D) used in order to describe the action in 4D space covering spatial coordinates, depth and time. Also represents the depth sequence in 4D grids by dividing the space and time axis into multiple segments. Another 4D descriptor proposed by called Random Occupancy Pattern (ROP) which deals with noise and occlusion combined with sparse coding approaches to increase robustness. Action recognition from different side views has been applied to gain more discriminative features. Generates side view from the front view of the depth map, both views are transformed to DMA (Depth Motion Appearance) descriptor and DMH (Depth Motion History) descriptor. Then, SVM is trained with the two descriptors to classify the action. Recently generate top and side views by rotating 3D points from the front view. The three views are used as inputs to three convolutional neural network models for feature extraction and action classification. In parallel to depth-based approaches, skeleton-based methods also have a huge contribution to the action recognition research area. In, each joint is associated with a Local Binary Pattern descriptor which is translation invariant and provide highly discriminative features. Additionally, a temporal motion representation called Fourier Temporal Pyramid is also proposed in order to model the joints movements. Eigen Joints is a new type of features proposed in to combine action information including static postures, motion and offset features. A framework based on sparse coding and temporal pyramid matching is proposed in for better 3D joint features representation. A histogram of 3D joint location called HOJ3D in represents the human joint's locations. Then, a posture words are built from HOJ3D vectors and trained using a Hidden Markov Model to classify the actions. In a framework is proposed for online human action recognition using a new Structured Channeling Skeletons feature (SSS) which can deal with intra-class variations including viewpoint, anthropometry, execution rate, and personal style. proposed non-parametric Moving Pose (MP) for low latency human action and activity recognition, the framework considers pose information, speed, and acceleration of the joints in the current frame within a time window. A hierarchical dynamic framework was reported in based on using deep belief networks for feature extraction and encoding dynamic structure into a HMM-based model. Addresses action recognition in videos by modelling the spatial-temporal structures of human poses. The method improves the pose estimation first, then groups the joints into five body parts. Moreover, data mining techniques have been applied to get spatial-temporal pose structures for action representation and transform the joint coordinates to a 2D image descriptor. A convolutional neural network model is used for action classification from the descriptor. Very recent works: SOS and Joint Trajectory Maps propose a new approach which transforms the skeleton joints trajectories shapes from 3D space into three images that represent the front view, the top view and the side view of the joints' trajectory shapes. Three convolutional neural networks extract features from the three images to classify the action. Convolutional neural network is a powerful technique for feature extraction and classification. Recent action recognition approaches started to focus more on using CNN for action classification rather than using SVM. Researchers in deep learning try always to come up with new techniques to improve the CNN architectures and enhance the performance of feature extraction, classification and computation speed. Summaries recent advances in convolutional neural network in term of regularization, optimization, Activation functions, loss functions, weight initialization and so on. Recent CNN based action recognition methods are based on using multiple action representations that employ many CNN channels for the processing. In many features concatenation architectures are proposed in order to improve the classification accuracy using multiple sources of knowledge. In spite of the fact that the previous approaches achieved good results, the problem of action recognition is still open and require more robust action representations and feature extraction techniques to improve the accuracy and overcome the weakness of the previously mentioned methods. To this end, the proposed work in this paper investigates the use of both types of data, depth maps and postures to enhance the action recognition throw the power of CNN for feature extraction and classification.

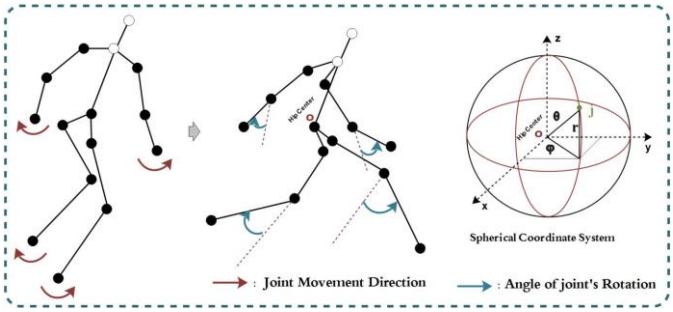

Figure 1. Human body joints motion direction during a running action. The joints motion more subject to a rotation, which makes the spherical coordinate system more suitable to represent the joints movements.

3. 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 2.

To start with, the color 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.

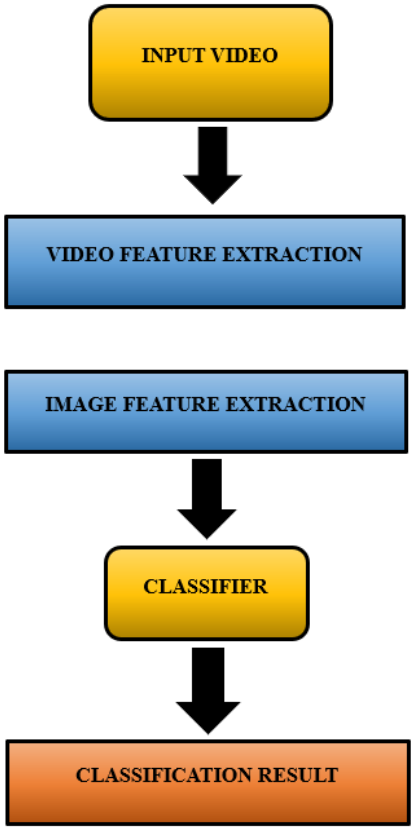

Figure 2. Flow chart of the proposed work

4. METHODLOGY

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. The color 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.

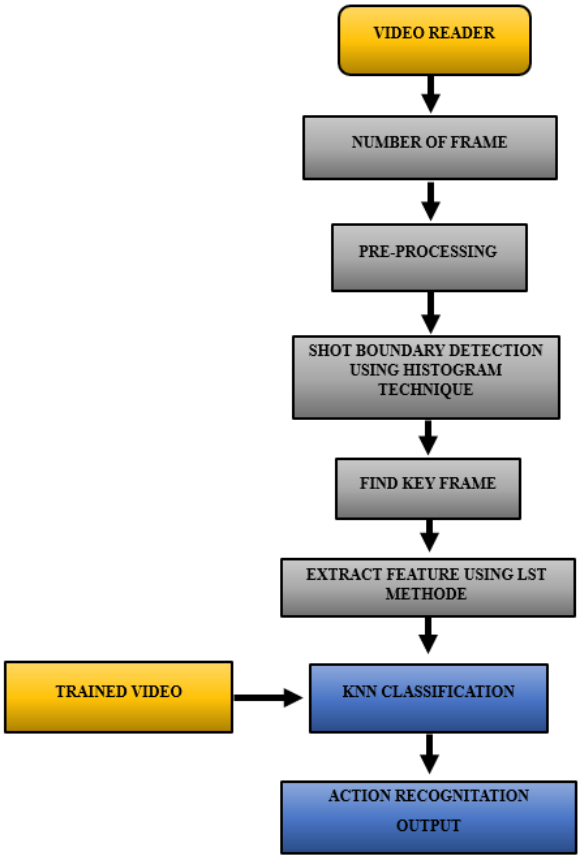

Figure 3. Process of the Action recognition system

5. IMPLIMENTATION, WORKING AND RESULT

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.

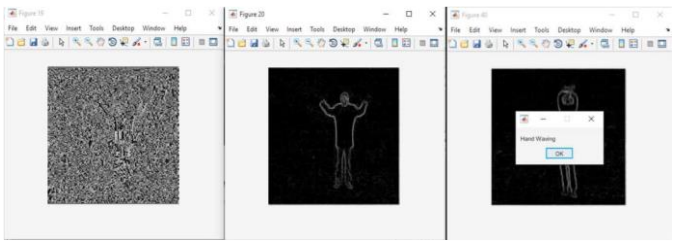

Figure 4. Test output for hand waving

The first image in figure 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.

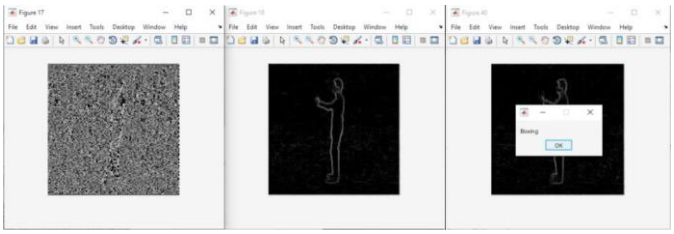

Figure 5. Test output for boxing

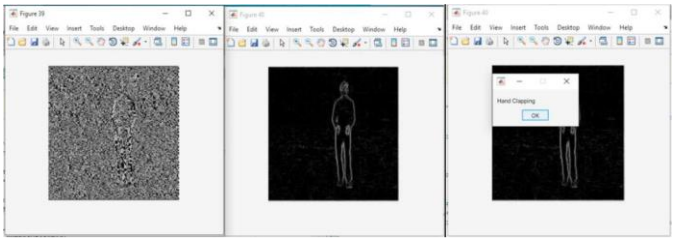

Figure 6. Test output for hand clapping

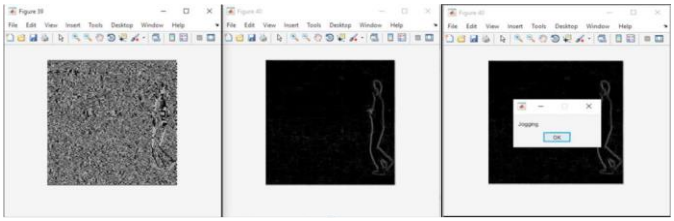

Figure 7. Test output for jogging

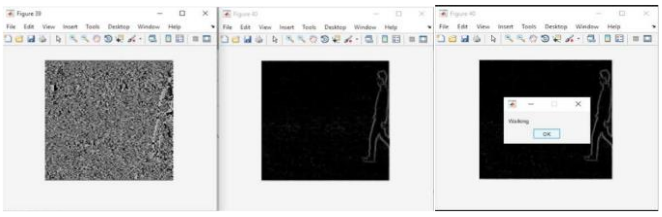

Figure 8. Test output for walking

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.

CONCLUSION

From above result presented in this paper we can conclude that a video action recognition system is proposed different hand and human body 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.

ACKNOWLEDGEMENT

We are deeply grateful to all those who contributed to the success of this review research paper. First and foremost, we would like to thank our primary supervisor Prof. Snehil G. Jaiswal, for their guidance, support, and encouragement throughout the entire process. Their mentorship and expertise were invaluable in helping us to shape the direction of our review research and to bring our ideas to fruition.

I would also like to thank the organizations and individuals who provided me a support for this review research, including **G.H. RAISONI University** Amravati, Maharashtra, India. Without their generous contributions, this review research would not have been possible.

Overall, this research project would not have been possible without the support and contributions of so many people. We are deeply grateful to all of those who helped to make this project a reality, and we hope that our findings will make a meaningful contribution to the field.

REFERENCES

1. H.Wang,A .Klaser,C.Schmid and C-L.Liu, “Action recognition by dense trajectories ,” in Proc. IEEE Conf. Comput. Vis.Pattern Recog., Jun. 2011, pp 3169-3176.
2. H.Wang and C Schmiid , “Action recognition with improved trajectories,” in Proc.IEEE Int. Conf.Comput. Vis., Dec 2013, pp 3551-3558.
3. Y-G Jiang,Q.Dai,X.Xue,W.Liu and C-W Ngo. “Trajectory-based modelling of human actions with motion reference points,” inProc. Eur.Conf .Comput.Vis.,Oct 2012,Vol 7576,pp.425-438.
4. M. D. Rodriguez, J. Ahmed, and M. Shah, “Action MACH: A spatiotemporal maximum average correlation height filter for action recognition,”in Proc. IEEE Conf. Comput. Vis. Pattern Recog., Jun. 2008, pp. 1–8.
5. Y. Hu, L. Cao, F. Lv, S. Yan, Y. Gong, and T. S. Huang, “Action detection in complex scenes with spatial and temporal ambiguities,” in Proc. IEEE Conf. Comput. Vis. Pattern Recog., Sep.–Oct. 2009, pp.128–135.
6. T. Lan, Y. Wang, and G. Mori, “Discriminative figure-centric models for joint action localization and recognition,” in Proc. IEEE Int. Conf. Comput. Vis., Nov. 2011, pp. 2003–210.
7. M.Raptis, I.Kokkinos and S.Soatto,” Discovering discriminative action parts from mid-level video representations” ,in Proc ,IEEE Conf.Comput.Vis,.Pattern Recog.,Jun 2012,pp.1242-1249
8. W. Brendel and S. Todorovic, “Learning spatiotemporal graphs of human activities,” in Proc. IEEE Int. Conf. Compute. Vis., Nov. 2011,
9. M.Jain,J.van Genmert,H.Jegou ,P.Bouthemy and C.Snoek ,”Action localization with tubelets from motion” ,in Proc IEEE Conf, Comput.Vis.Pattern Recog. Jun 2014 pp 740-747
10. 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.
11. Ronald Poppe, “A survey on vision-based human action recognition”, The Netherlands Image and Vision Computing, vol. 28 (2010), Pp. 976–990.
12. 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.

13. 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.
14. 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.
15. D.M. Gavrila and L.S. Davis "3D model-based tracking of humans in action: a multi-view approach", IEEE. Pp. 73-80.
16. 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.
17. 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.
18. 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.
19. 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.
20. 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,
21. 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.

BIOGRAPHIES

I'm pursuing my Masters of Technology degree from **G. H. RAISONI** University Amravati, Maharashtra, India. This paper is our academic research paper submitted and created to Recognize Human Detection from any Video.

NAME: Mr. Dhananjay Arvindrao Deshpande
Email: deshpandedhananjay84@gmail.com
Address: Flat no 102, Classic view Apartment, Venkatesh Colony, Wadali, Camp, Amravati, Maharashtra