A Comprehensive Survey on Human Action Recognition

Mallikarjun Aralimarad, Meena S M, Jayashree D Mallapur

Abstract: *The present The present situation is having many challenges in security and surveillance of Human Action recognition (HAR). HAR has many fields and many techniques to provide modern and technical action implementation. We have studied multiple parameters and techniques used in HAR. We have come out with a list of outcomes and drawbacks of each technique present in different researches. This paper presents the survey on the complete process of recognition of human activity and provides survey on different Motion History Imaging (MHI) methods, model based, multiview and multiple feature extraction based recognition methods.*

Keywords : *Computer Vision, HAR, , Histogram of Oriented Gradients(HOG), MHI.*

I. INTRODUCTION

In the present days Human Action Recognition (HAR) is an emerging area of research field of computer vision. The terms “action” and “activity” are oftenly used in the vision literature and are used as similar words [1, 2]. In the following discussion, by “actions” are very discrete and they constitute simple motion patterns and are mostly done by a solo human being and naturally lasting for small amount of time, maximum lasting for ten of seconds, running, walking, etc are the examples for actions. In another event, the meaning of “activities” is compound chain of actions performed by group of humans who could be making conversation or doing a task with each other in a controlled manner. Remarkable quantities of videos are recorded in a day due to the closed circuit TV cameras, cinema, serials, musical.ly, etc. The aim of HAR is to find the behavior of human base on the actions performed by him and intention of one or more objects. But to achieve the goal of finding the action or activities in a given scenario or environment is always challenged by adversities.

Manuscript received on May 25, 2020.
Revised Manuscript received on June 29, 2020.
Manuscript published on July 30, 2020.
* Correspondence Author

Mallikarjun Aralimarad, Assistant Professor of Electronics and Communication Engineering, Basaveshwara Engineering College (Autonomous), Bagalkot, Karnataka, India.

Dr. Meena S. M., HOD and Professor School of Computer Science and Engineering, KLETECH, Hubli. India.

Dr. Jayashree D. Mallapur, Professor Department of Electrictronics and Communication Engineering department at Basaveshwar Engineering College (Autonomous), Bagalkot. India

© The Authors. Published by Blue Eyes Intelligence Engineering and Sciences Publication (BEIESP). This is an open access article under the CC BY-NC-ND license (http://creativecommons.org/licenses/by-nc-nd/4.0/)

HAR depends mainly on the context and environment where it is being performed [3]. It has several levels of complexity such as body parts motion (gestures and sign languages), movements (walking, running), actions(reaching and following), Human-Object interaction(grasping and punching), Human-Human interaction(handshaking and punching), and social behavior (leading and emphasizing).

In general, the recognition of human activity from video consist four different steps, shown in below block diagram Fig.1.

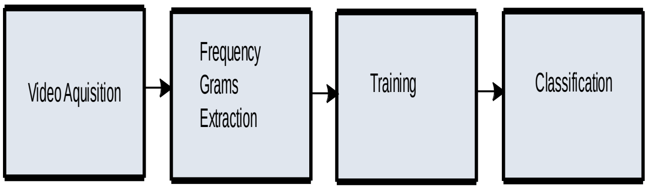

Fig.1. Block diagram of HAR.

(I) Acquisition of input video and Pre-Processing (extraction of image or frames, filtering for remove noise etc, segmentation of ROI). This is a very important step which makes the feature extraction simple task. **(II)** Feature extraction is performed by motion tracking, in which basic idea is to detect moving object in frame and extract the features like data of pose, joint point of skeleton, blob trajectory, histogram etc. **(III)** Training given to algorithm using feature vector obtained from different features to classify different movements. This algorithm may be supervised or unsupervised based on the scenario. **(IV)** Finally, Classification is performed by classifier to identify and investigate the activities from the videos. The general structure of HAR system is depicted in Fig.1. There are so several data sets available for HAR. Some of the popular data sets are MSR Action 3-D Dataset, Florence 3-D Action Dataset, UTKinect Dataset, Kinect data sets, Berkeley MHAD dataset, NTU RGB+D dataset, UCF-100 dataset, etc [57]. The paper is organized as follows. Section 2 mentions about the various Issues in HAR emerged out of literature survey. Section 3, 4, 5, 6 discusses about the motion, multifeature, multiview and model based techniques for HAR respectively. Section 7 discusses about the challenges and future research directions followed by conclusion in Section 8. Several papers are published on HAR based on Spatio-Temporal Interest Point (STIP) Detection Algorithms [3, 6], video based [1, 2, 3, 5, 7, 8, 40, 65, 66, 70], analysis techniques [9], multiview [39], Handcrafted and deep learning [4, 15]. The emerged issues are as follows:

- The camera angle plays a very important role in HAR. Because in real world camera angle is not constant and keeps on changing so, the performance of HAR system should not be altered as the camera angle changes.

- HAR system should not be sensitive to the changes in lighting conditions and occlusion in a given frame.
- Very low intraclass differences and more interclass difference always create problems in HAR system
- Appearance of human changes because of the way of performing activities continues changing dependent on the plane on which activity is performed, dress likewise assume an imperative job in the presence of human and the items they convey with them.

Hence it is been an exploration issue to perceive human activity wrong to the presence of human.

- The actions performed by humans keeps on varying based on the context or plane on which it is done. The dress also changes the human appearance and the objects which are attached with it. So this is also a very important research topic in HAR systems.
- Recognizing the exercises of human with a jumbled or dynamic foundation is troublesome [3]. The nature of a video is additionally a main consideration in choosing the execution of the framework. A productive movement acknowledgment framework ought to perceive the human even in the differing prevalence of the video and the jumbled foundation.
- In a given video background of the video is very important in deciding the performance of HAR system [3]. Because the background may cluttered or full of activities. The quality of video also plays very important role and a HAR system should be able to identify the activity in any given quality video.
- The principle difficulties of deep learning are [15] adaptability of calculations. Also, there are some different difficulties, for example, the requirement for a specialist to structure a suitable system because of various hyper parameters, demonstrating a few components of varieties and the connections between them, large number of nearby minima, execution of Stochastic Gradient Descent (SGD) in numerous layers, over-fitting issue , and the require a lot of information.
- Nowadays the deep learning methods are used in HAR system [15]. Deep learning methods also face several challenges like and lack of appropriate illustration learning measure, disentangling fundamental factors of variation and non-convex optimization. Other than mentioned challenges they also have over fitting problem, several hyper parameters selection, considering several factors which are varying and they interacting with each other, handling large number of local minima. They also need a huge amount of data for training to improve the performance.

The whole literature survey has been classified into four categories of analysis, which are as shown in Fig.2

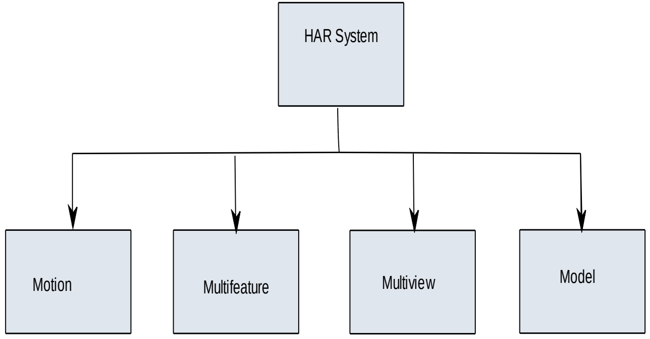

Fig.2. Classification HAR Techniques

II. MOTION BASED TECHNIQUES FOR HAR

This review discusses about the recognition based on motion energy history [38]. HAR framework dependent on Multi-Velocity Spatio-Temporal Interest Points (MVSTIPs) and a novel neighborhood descriptor called Motion Energy (ME) Orientation Histogram (MEOH) [18] is introduced which gives high precise outcomes. The body variation is taken as a key point and this is discriminated by the adaptive boosting algorithm , which helps to recognize the human action[10]. By using the neural networks along with skeleton, which is in irregular shape is again structured to get unidirectional attribute graph is presented in [19]. A hybrid system with fast Histogram of Oriented Gradients(HOG) 3D of video recordings and Smith–Waterman incomplete shape coordinating are blended and Non-linear Support Vector Machine (SVM) decision trees characterize and discover the activities in a given video [20]. A 3D skeleton kinematic joint model uses less computational time and this element is utilized in a HAR framework to get elite for pragmatic utilization [21]. A quick, straightforward, and intense technique for HAR was proposed which is dependent on human kinematic similarity. The angular characteristics are used which provide the simple normalization. 3 k-Nearest Neighbors (3k-NN) are used one for training, one to find the frame label and last one for classification [22]. Depth Motion Maps (DMMs) are additionally used to show signs of improved HAR results. In the proposed method distance weighted Tikhonov matrix is used for the action recognition [23, 68, 69]. The HAR system is too much dependent on environment and is most of the time fixed and handcrafted features. This becomes limitation for HAR. To solve the problem Skeleton Motion History Images (SMHI) of human activities is utilized to realize HAR system that can work autonomously on the issue. The SMHI features are used for deep learning which provides excellent results even in real time. A variation is done where the Long Short-Term Memory (LSTM) is used with spatial and temporal skeletal data for HAR which also gives the excellent results [24, 14]. The 3-D HAR is also a emerging field in research. The challenge of HAR based on the shape of motion trajectories on Riemannian Manifold [46]. DMMs features are found front, top and side and Local Binary Patterns (LBPs) are used to get compact representation of features [47]. Motion characteristics can be mentioned in terms of skeletal joints. The human skeletal is divided into five parts and given local feature extractor and fused to get HAR [48]. The LSTM is used in HAR but the success is not that good when it comes to the similarity in the actions. So a novel frame work has been realized based on Shape Evolution Maps (SEM) along with Motion Evolution Maps (MEM) [49]. The skeletal based HAR techniques always face problem similarity in the postures which is a challenging problem. This problem is addressed in a framework which realizes a new Relation Matrix of 3D Rigid Bodies (RMRB3D) which is compact representation of poses [58]. In another HAR framework a sequence of skeletal features are used as features and the histograms of these features is derived [59]. A fastmap based HAR framework was proposed which uses short temporal set of fastmap to represent raw moving silhouettes. The method used very simple and fast [60].

III. MULTIFEATURE BASED TECHNIQUES FOR HAR

This literature discusses about HAR techniques which use multifetures. A hybrid feature technique is proposed and implemented for which has two stages feature extraction and recognition [12, 25].

A novel system has been planned and realized where numerous features for activity recognition are combined top to depth sequence. Two sorts of features separated: I) a quantized vocabulary of neighborhood spatio-temporal descriptor HOG3D, and ii) a global projection based descriptor that figures the HOG from the DMMs [26]. A novel structure incorporating modules with key edges extraction by meager requirement and afterward the combination multi-include was built and Max-pooling strategy individually [27]. Two sets of features of RGBD recordings are used in HAR [42]. In a proposed model extricating sets of spatial and temporal local features from subgroups of joints are used, which are totaled by a strong strategy. A few element vectors are blended by a metric learning strategy propelled by the LMNN algorithm with an aim to build the performance utilizing the nonparametric k-NN classifier [28]. The frequency grams which are nothing but the features derived based on histograms of the motion along with its spatiotemporal gradient are used in HAR [50]. Spatio-temporal features are derived then these are described by HOG-HOF descriptors and fused reduced in HAR which gave the excellent results [51]. In a frame the Human action region is found using frame difference and VIBE algorithm. Then the (HOG3D) extracted and then STIPs are extracted. The extracted features are given to PCA to reduce the dimensionality [52]. A trained model is used for feature extraction and then followed by SVM and KNN for classification of actions [72] The RGB features along with depth maps are used in HAR. But face difficulty in fusing the two features. A novel method is proposed in [74] using bag of visual words and multi class support vector machine. We have features like LBP, HOG, HAR wavelets, velocity and displacement. By fusing these features to get maximum efficiency is done in [75].

IV. MULTIVIEW BASED TECHNIQUES FOR HAR

The section mainly discusses about the HAR techniques which use multiview in a given data or scenario. Unsupervised feature fusion technique named as Multiview Cauchy Estimator Feature Embedding (MCEFE)) is utilized in a novel HAR framework [29]. Identifying the human actions from an unknown view or unseen view is a challenging task. A Robust Non-Linear Knowledge Transfer Model (R-NKTM) model was proposed and realized which gives excellent results for novel views [30]. A Fully-associated Neural Network (NN) is utilized that exchanges information of human activities to an obscure view to a mutual abnormal state virtual view by finding an arrangement of non-straight changes that interfaces the perspectives [31]. Multiview image sequences approach is used in novel HAR system which is based on finding local motion from multiple camera angles. In place of MHI the HOG features are used and are classified using k-NN. The proposed model has the advantage of less memory and bandwidth requirement and also computationally very efficient which makes it suitable for real time scenarios [32]. The multiview approach is used in layer fusion model where the fusion of multilayer features takes which in turn used for action recognition [43]. Simple multiview based action recognition is implemented using combined scale invariant contour-based pose features from silhouettes and uniform rotation invariant LBP are extracted [44]. In another approach the of person-person interaction two different features from different view are fused and used for action recognition [45]. Multitask structural learning (MTSL) problem is used in HAR framework where multiview/single view HAR is realized. The MTSL has a advantage of preserving consistence between the body based and action based classification and also discovers the action-specific and action-shared feature subspaces which in turn strengthen the model learning [53]. The view-invariant features are used in HAR framework to describe the different action in multiviews. The features are derived using holistic features like temporal points of interest and are going to represent the global spatio temporal features [54]. In recognizing the action in a video the challenge is always the variation in view point. The view point variation is alleviated by neural networks using RNN and LSTM in [67].

V. MODEL BASED TECHNIQUES FOR HAR

In an HAR system the recognition can be done using sequence of frames or single frame. A You Only Look Once (YOLO) approach was introduced to find, localize and identify actions of most interest. The frames were obtained from a surveillance camera. It is interesting to know that the YOLO is based on CNN which is applied to complete image and modifies the image into grids and these grids provide the required features [33]. A Chinese Restaurant Process is utilized in HAR to explicitly describe these one of a kind inner models of a specific complex movement [34]. Long-term temporal convolutions models are utilized and expanded temporal degrees have enhanced the exactness of activity acknowledgment [35]. A minimal effort descriptor called 3D Histograms of Textures (3DHoTs) to remove discriminant highlights from a succession of profundity maps. 3DHoTs are derived from anticipating depth scenes onto three orthogonal Cartesian planes, i.e., the frontal, side, and best planes, and accordingly minimalistically portray the notable data of a particular activity, on which surface highlights are computed to speak to the activity [36]. A superior comprehension of human collaborations in recordings can be accomplished by all the while thinking about the coarse communications between individuals, the activity of every person, and the action surprisingly in general [13]. Gaze assisted deep neural system, which plays out the activity acknowledgment undertaking with the assistance of human visual consideration, has been introduced. By using the pooling concept and dynamic gaze concept the proposed model provides excellent results [17]. In the recent years Recurrent Neural Network (RNN) has gained lot of popularity and framework has been proposed for HAR based on RNN. In this the spatio-temporal attention feature mechanism is used [16, 71].

A novel Improved Gaussian Process Latent Variable Model (GPLVM) used to for feature dimensionality reduction and Hidden Conditional Random Field (HCRF) for action recognition which gives a accuracy of 93.68% [41].

In another method CRFs used to recognize activity by considering the parameters through SVM [37]. The rotational features are derived using R-transform and Spatial Distribution of Gradients (SDGs), Sum of Directional Pixels (SDPs) are derived from Average Energy Silhouettes Images (AESIs) to describe the shape and are used for the recognition [42]. Fuzzy models like Type 2 fuzzy sets along with message passing algorithms is used for HAR[11]. In a novel HAR framework the Bag of Words (BoW) approach is used. The BoWs models are based on the covariance matrices of spatio-temporal features and these features are derived from histogram of optical flow. The Symmetric Positive Definite (SPD) matrix, along with the log-Euclidean geometry is derived for distinguishing between covariance matrices [55]. A novel framework was proposed in which a probability based HAR system realized. The spatial-temporal action features and scene descriptors are extracted naive Bayes nearest neighbour algorithm is used for classification [56]. The knowledge–information–data (KID) model is utilized for learning cooperative information while the AMR ceaselessly searches for relationship among information units and consolidations related units utilizing blending instruments [61]. In certain exercises the movements are comparable, so to determine the issue a Discriminative Group Context Features (DGCF) that considers unmistakable sub-occasions. Also, we embrace a Gated Recurrent Unit (GRU) model that can learn worldly changes in a succession. In true situations, individuals perform exercises with various fleeting lengths [62]. Instead of atomic actions which are simple complex actions were difficult to recognise. Complex activity recognition is done by Bayesian networks along with GPA are used in [63]. Multi-modality principal orientations and residual descriptors method by reducing the modality gap of the RGB channels and the depth channel at the feature level to make our method applicable to RGB-D action recognition [64, 73].

Table- II: Name of the Table that justify the values

Literature Listing	Technique adopted	Disadvantages
[10][14][18][19][20][21][22][23][24][38][46][47][48][49][58][59][60][68][69][72]	Motion in given video.	Homogeneous actions have same motion which is still not separated by these techniques.
[12][25][26][27][28][50][51][52][74][75]	Multifeature	Many features are fused to get the best efficiency but redundancy in data increased. So the processing time.
[29][30][31][32][43][44][45][53][54][67]	Multiview	Always it is difficult to find optimum orientation for different viewpoints.
[11][13][33][34][35][36][37][16][17][41][42][55][56][71][61][62][63][64][65][73]	Many model like Bayesian, CNN, RNN, LSTM etc..	Even with the different models used for recognition the efficiency varies with different action and data sets.

VI. CHALLENGES AND FUTURE RESEARCH DIRECTIONS

HAR has several issues which are to be addressed more aggressively and they deserve more research. Some challenges are discussed in this section. In a given scene if small actions like twigs of tree and bushes blowing in air create a noisy image or video pose threat to performance of the algorithm.. Gradual lighting conditions in given scene and meager quality of image also pose threat to performance index. Several objects moving in a scene may be for long or short duration and shadows of the objects also create problem in HAR. Camouflage, dress and change in style of human also create problems in HAR system. Several actions have similarity and it’s still an open research topic in HAR. Tracking multiple objects and finding missing limb or joint is still a challenge. Finding abnormal behavior of crowd is still an open research topic. In the above discussed many papers some have provided excellent solutions but they are still far away from real world timely constraints. The previous research discusses much about the topic of action classification rather than action detection.

VII. CONCLUSIONS

There are varieties of HAR algorithms and also several methodologies are present, but our survey shows majority of work done in recent years using MHI, multifeture technique, multiview technique and different models used in HAR system. This paper is to review the previous works that are done in HAR. We have classified the techniques used in HAR and accounted their benefits. In each of the classifications we had challenges that are not completely resolved to have the smooth conduction of HAR. The following are the challenges which we want to address: extraction of low level features for abnormal HAR in the influence of noisy data, Video quality affected by shadows, occlusion, illumination, moving camera, and complex backgrounds and spatiotemporal variations for the same activity. Multiple actions of motion can be separated, analyzed by the techniques which already exist.

But present papers have not taken homogeneous action for considerations.

Multifeature techniques are leading to redundancy of data extraction and in turn increasing the processing time and also difficulty in fusing the multifeature into one framework. There is always optimization challenge for multiview, because in HAR there is always scope for multiview and multi angle consideration. There are many models have been proposed but they not much of work done in fusion model with many features. The models presented have not optimized there solution for multiple features. The result of the research shows the solution to challenges discussed so far can be answered by deep neural networks.

REFERENCES

1. P. Turaga, R. Chellappa, V. S. Subrahmanian, and O. Udrea, “Machine recognition of human activities: A survey”,2008, *IEEE Transactions on Circuits and Systems for Video technology*, 2008, vol. 18, no. 11, p. 1473.
2. R. Poppe, “A survey on vision-based human action recognition”, 2010, *Image Vis. Comput.*, vol. 28, no. 6, pp. 976–990, 2010, doi: 10.1016/j.imavis.2009.11.014
3. Bhoomika Rathod, Devang Pandya,Raunakraj Patel, “A Survey on Human Activity Analysis Techniques”, International Journal on Future Revolution in Computer Science & Communication Engineering Volume: 3 Issue: 11 ISSN: 2454-4248 462 – 471.
4. A. B. Sargano, P. Angelov, and Z. Habib, “A comprehensive review on handcrafted and learning-based action representation approaches for human activity recognition”, 2017, *Appl. Sci.*, vol. 7, no. 1, 2017, doi: 10.3390/app7010110.
5. S. Zhang, Z. Wei, J. Nie, L. Huang, S. Wang, and Z. Li, “A Review on Human Activity Recognition Using Vision-Based Method”, 2017, *J. Healthc. Eng.*, vol. 2017, 2017, doi: 10.1155/2017/3090343.
6. Y. Li, R. Xia, Q. Huang, W. Xie, and X. Li, “Survey of Spatio-Temporal Interest Point Detection Algorithms in Video”, 2017, vol. 5, pp. 10323–10331, 2017.
7. C. J. Dhamsania and T. V. Ratanpara, “A survey on Human action recognition from videos”, in *Green Engineering and Technologies (IC-GET)*, 2016, *Online International Conference on*pp. 1–5.
8. T. Subetha and S. Chitrakala, “A Survey on human activity recognition from videos”, 2016, in *Information Communication and Embedded Systems (ICICES)*, 2016, *International Conference on* pp. 1–7.
9. X. Xu, J. Tang, X. Zhang, X. Liu, H. Zhang, and Y. Qiu, “Exploring techniques for vision based human activity recognition: Methods, systems, and evaluation”, 2013, *Sensors*, vol. 13, no. 2, pp. 1635–1650.
10. N. Zerrouki, F. Harrou, Y. Sun, and A. Houacine, “Vision-Based Human Action Classification Using Adaptive Boosting Algorithm” *IEEE Sens. J.*, vol. 18, no. 12, pp. 5115–5121, 2018, doi: 10.1109/JSEN.2018.2830743.
11. X.-Q. Cao and Z.-Q. Liu ,”Type-2 fuzzy topic models for human action recognition”, 2015, *IEEE Transactions on Fuzzy Systems*, vol. 23, no. 5, pp. 1581–1593.
12. M. Selmi, M. A. El-Yacoubi, and B. Dorizzi, “Two-layer discriminative model for human activity recognition”, 2016, *IET Computer Vision*, vol. 10, no. 4, pp. 273–279.
13. X. Zhen, F. Zheng, L. Shao, X. Cao, and D. Xu, “Supervised local descriptor learning for human action recognition”, 2017, *IEEE Transactions on Multimedia*, vol. 19, no. 9, pp. 2056–2065.
14. S. Song, C. Lan, J. Xing, W. Zeng, and J. Liu, “Spatio-Temporal Attention-Based LSTM Networks for 3D Action Recognition and Detection”, 2018, *IEEE Transactions on Image Processing*, vol. 27, no. 7, pp. 3459–3471.
15. M. Koohzadi and N. M. Charkari, “Survey on deep learning methods in human action recognition”, 2017, *IET Computer Vision*, vol. 11, no. 8, pp. 623–632.
16. W. Du, Y. Wang, and Y. Qiao, “Recurrent spatial-temporal attention network for action recognition in videos”, 2018, *IEEE Transactions on Image Processing*, vol. 27, no. 3, pp. 1347–1360.
17. Y. Liu, Q. Wu, L. Tang, and H. Shi, “Gaze-assisted multi-stream deep neural network for action recognition”, 2017, *IEEE Access*, vol. 5, pp. 19432–19441.
18. C. Li, B. Su, J. Wang, H. Wang, and Q. Zhang, “Human Action Recognition Using Multi-Velocity STIPs and Motion Energy Orientation Histogram”, 2014, *J. Inf. Sci. Eng.*, vol. 30, no. 2, pp. 295–312.
19. C. Li, Z. Cui, W. Zheng, C. Xu, R. Ji, and J. Yang, “Action-Attending Graphic Neural Network”, 2018, *IEEE Transactions on Image Processing*, vol. 27, no. 7, pp. 3657–3670.
20. Ibrahim El-Henawy, Kareem Ahmed, Hamdi Mahmoud, “Action recognition using fast HOG3D of integral videos and Smith–Waterman partial matching”, 2018, ET Image Process, 2018, Vol. 12 Iss. 6, pp. 896-908 © The Institution of Engineering and Technology.
21. J. Li, X. Mao, X. Wu, and X. Liang, “Human action recognition based on tensor shape descriptor”, 2016, *IET Computer Vision*, vol. 10, no. 8, pp. 905–911.
22. Q. Wu, G. Xu, L. Chen, A. Luo, and S. Zhang, “Human action recognition based on kinematic similarity in real time”, 2017, *PloS one*, vol. 12, no. 10, p. e0185719.
23. C. Chen, K. Liu, and N. Kehtarnavaz, “Real-time human action recognition based on depth motion maps”, 2016, *Journal of real-time image processing*, vol. 12, no. 1, pp. 155–163.
24. C. N. Phyo, T. T. Zin, and P. Tin, “Skeleton motion history based human action recognition using deep learning”, 2017, in *Consumer Electronics (GCCE), 2017 IEEE 6th Global Conference on* pp. 1–2.
25. S. Zhu and L. Xia, “Human action recognition based on fusion features extraction of adaptive background subtraction and optical flow model”, 2015, *Mathematical Problems in Engineering*, vol. 2015.
26. Q. Xiao and J. Cheng, “ Human action recognition framework by fusing multiple features”, 2013, in *Information and Automation (ICIA), 2013 IEEE International Conference on* pp. 985–990.
27. J. Li, X. Mao, L. Chen, and L. Wang, “Human interaction recognition fusing multiple features of depth sequences”, 2017, *IET Computer Vision*, vol. 11, no. 7, pp. 560–566.
28. D. C. Luvizon, H. Tabia, and D. Picard, “Learning features combination for human action recognition from skeleton sequences”, 2017, *Pattern Recognition Letters*, vol. 99, pp. 13–20.
29. Y. Guo, D. Tao, W. Liu, and J. Cheng, “Multiview Cauchy Estimator Feature Embedding for Depth and Inertial Sensor-Based Human Action Recognition”, 2017, *IEEE Trans. Systems, Man, and Cybernetics: Systems*, vol. 47, no. 4, pp. 617–627.
30. H. Rahmani, A. Mian, and M. Shah, “Learning a deep model for human action recognition from novel viewpoints”, 2018, *IEEE transactions on pattern analysis and machine intelligence*, vol. 40, no. 3, pp. 667–681.
31. S. Chun and C.-S. Lee, “Human action recognition using histogram of motion intensity and direction from multiple views”, 2016, *IET Computer vision*, vol. 10, no. 4, pp. 250–257.
32. F. Murtaza, M. H. Yousaf, and S. A. Velastin, “Multi-view human action recognition using 2D motion templates based on MHIs and their HOG description”, 2016, *IET Computer Vision*, vol. 10, no. 7, pp. 758–767.
33. S. Shinde, A. Kothari, and V. Gupta, “ YOLO based Human Action Recognition and Localization”, 2018, *Procedia computer science*, vol. 133, pp. 831–838.
34. L. Liu, L. Cheng, Y. Liu, Y. Jia, and D. S. Rosenblum, “Recognizing Complex Activities by a Probabilistic Interval-Based Model”, 2016, in *AAAI*, 2016, vol. 30, pp. 1266–1272
35. G. Varol, I. Laptev, and C. Schmid, “Long-term temporal convolutions for action recognition”, 2018, *IEEE transactions on pattern analysis and machine intelligence*, vol. 40, no. 6, pp. 1510–1517.
36. B. Zhang, Y. Yang, C. Chen, L. Yang, J. Han, and L. Shao, “Action recognition using 3D histograms of texture and a multi-class boosting classifier”, 2017, *IEEE Trans. Image Process*, vol. 26, no. 10, pp. 4648–4660.
37. Z. Wang, S. Liu, J. Zhang, S. Chen, and Q. Guan, “A Spatio-Temporal CRF for Human Interaction Understanding” , 2017, *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 27, no. 8, pp. 1647–1660.
38. J. K. Aggarwal and M. S. Ryoo, “ Human activity analysis: A review” , 2011, *ACM Computing Surveys (CSUR)*, vol. 43, no. 3, p. 16.
39. M. B. Holte, C. Tran, M. M. Trivedi, and T. B. Moeslund, “Human action recognition using multiple views: a comparative perspective on recent developments” , 2011, in *Proceedings of the 2011 joint ACM workshop on Human gesture and behavior understanding*, 2011, pp. 47–52.
40. L. Cai, X. Liu, H. Ding, and F. Chen, “Human Action Recognition Using Improved Sparse Gaussian Process Latent Variable Model and Hidden Conditional Random Filed” , 2018, *IEEE Access*, vol. 6, pp. 20047–20057.

41. D. K. Vishwakarma and K. Singh, "Human Activity Recognition Based on Spatial Distribution of Gradients at Sublevels of Average Energy Silhouette Images" , 2017, *IEEE Transactions on Cognitive and Developmental Systems*, vol. 9, no. 4, pp. 316–327.
42. Rawya Al-Akam, Dietrich Paulus, "RGBD Human Action Recognition using Multi-Features Combination and K-Nearest Neighbors Classification" , 2017, International Journal of Advanced Computer Science and Applications, Vol. 8, No. 10.
43. P. Chalearnnetkul and N. Suvonvorn, "Multiview Layer Fusion Model for Action Recognition Using RGBD Images" , 2018, *Computational Intelligence and Neuroscience*, vol. 2018.
44. A. K. S. Kushwaha, S. Srivastava, and R. Srivastava, "Multi-view human activity recognition based on silhouette and uniform rotation invariant local binary patterns" , 2017, *Multimedia Systems*, vol. 23, no. 4, pp. 451–467.
45. M. Li and H. Leung, "Multi-view depth-based pairwise feature learning for person-person interaction recognition" , 2017, *Multimedia Tools and Applications*, pp. 1–19.
46. M. Devanne, H. Wannous, S. Berretti, P. Pala, M. Daoudi, and A. Del Bimbo, "3-d human action recognition by shape analysis of motion trajectories on riemannian manifold" , 2015, *IEEE transactions on cybernetics*, vol. 45, no. 7, pp. 1340–1352.
47. C. Chen, R. Jafari, and N. Kehtarnavaz, "Action recognition from depth sequences using depth motion maps-based local binary patterns", 2015, in *Applications of Computer Vision (WACV), 2015 IEEE Winter Conference on*, 2015, pp. 1092–1099.
48. Y. Du, Y. Fu, and L. Wang, "Representation learning of temporal dynamics for skeleton-based action recognition" , 2016, *IEEE Transactions on Image Processing*, vol. 25, no. 7, pp. 3010–3022.
49. H. Liu, J. Tu, M. Liu, and R. Ding, "Learning Explicit Shape and Motion Evolution Maps for Skeleton-Based Human Action Recognition" , 2018, in *2018 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, 2018, pp. 1333–1337.
50. T. Sandhan and J. Y. Choi, "Frequencygrams and multi-feature joint sparse representation for action and gesture recognition" , 2014, in *Image Processing (ICIP), 2014 IEEE International Conference on*, 2014, pp. 1450–1454.
51. A. I. Maqueda, A. Ruano, C. R. del-Blanco, P. Carballeira, F. Jaureguizar, and N. García, "Novel multi-feature bag-of-words descriptor via subspace random projection for efficient human-action recognition" , 2015, in *Advanced Video and Signal Based Surveillance (AVSS), 2015 12th IEEE International Conference on*, 2015, pp. 1–6.
52. W. Song, N. Liu, G. Yang, F. Lin, and P. Yang, "Multi-feature fusion based human action recognition algorithm" , 2015.
53. A.-A. Liu, Y.-T. Su, P.-P. Jia, Z. Gao, T. Hao, and Z.-X. Yang, "Multiple/single-view human action recognition via part-induced multitask structural learning" , 2015, *IEEE transactions on cybernetics*, vol. 45, no. 6, pp. 1194–1208.
54. K.-P. Chou *et al,* "Robust Feature-Based Automated Multi-View Human Action Recognition System" , 2018, *IEEE Access*, vol. 6, pp. 15283–15296.
55. M. Faraki, M. Palhang, and C. Sanderson, "Log-Euclidean bag of words for human action recognition" , 2014, *IET Computer Vision*, vol. 9, no. 3, pp. 331–339.
56. H.-B. Zhang *et al.*, ""Probability-based method for boosting human action recognition using scene context" , 2016, *IET Computer Vision*, vol. 10, no. 6, pp. 528–536.
57. J. M. Chaquet, E. J. Carmona, and A. Fernández-Caballero, "A survey of video datasets for human action and activity recognition" , 2013, *Computer Vision and Image Understanding*, vol. 117, no. 6, pp. 633–659.
58. W. Ding, K. Liu, G. Li, and X. Ran, "Human action recognition using spectral embedding to similarity degree between postures" , 2016, in *Visual Communications and Image Processing (VCIP), 2016*, pp. 1–4.
59. E. Cippitelli, E. Gambi, S. Spinsante, and F. Florez-Revuelta, "Human Action Recognition Based on Temporal Pyramid of Key Poses Using RGB-D Sensors" , 2016, in *International Conference on Advanced Concepts for Intelligent Vision Systems*, 2016, pp. 510–521.
60. L. C. Belhadj and M. Mignotte, "Spatio-temporal fastmap-based mapping for human action recognition" , 2016, in *Image Processing (ICIP), 2016 IEEE International Conference on*, 2016, pp. 3046–3050.
61. R. Huang, P. K. Mungai, J. Ma, I. Kevin, and K. Wang, "Associative memory and recall model with KID model for human activity recognition" , 2019, in *Future Generation Computer Systems*, vol. 92, pp. 312–323.
62. P.-S. Kim, D.-G. Lee, and S.-W. Lee, "Discriminative context learning with gated recurrent unit for group activity recognition" , 2018, in *Pattern Recognition*, vol. 76, pp. 149–161.
63. L. Liu, S. Wang, B. Hu, Q. Qiong, J. Wen, and D. S. Rosenblum, "Learning structures of interval-based Bayesian networks in probabilistic generative model for human complex activity recognition" , 2018, *Pattern Recognition*, vol. 81, pp. 545–561.
64. L. Chen, Z. Song, J. Lu, and J. Zhou, "Learning principal orientations and residual descriptor for action recognition" , 2019, in *Pattern Recognition*, vol. 86, pp. 14–26.
65. I. Jegham, A. B. Khalifa, I. Alouani, and M. A. Mahjoub, "Vision-based human action recognition: An overview and real world challenges" , 2020, in *Forensic Science International: Digital Investigation*, vol. 32, p. 20090.
66. C. Dhiman and D. K. Vishwakarma, "A review of state-of-the-art techniques for abnormal human activity recognition" , 2019, in *Engineering Applications of Artificial Intelligence*, vol. 77, pp. 21–45.
67. P. Zhang, C. Lan, J. Xing, W. Zeng, J. Xue, and N. Zheng, "View adaptive neural networks for high performance skeleton-based human action recognition" , 2019, in *IEEE transactions on pattern analysis and machine intelligence*, vol. 41, no. 8, pp. 1963–1978.
68. K. Kim, A. Jalal, and M. Mahmood, "Vision-Based Human Activity Recognition System Using Depth Silhouettes: A Smart Home System for Monitoring the Residents" , 2019, in *Journal of Electrical Engineering & Technology*, vol. 14, no. 6, pp. 2567–2573.
69. A. Jalal, S. Kamal, and C. A. Azurdia-Meza, "Depth maps-based human segmentation and action recognition using full-body plus body color cues via recognizer engine" , 2019, in *Journal of Electrical Engineering & Technology*, vol. 14, no. 1, pp. 455–461.
70. H.-B. Zhang *et al.*, "A comprehensive survey of vision-based human action recognition methods" , 2019, in *Sensors*, vol. 19, no. 5, p. 1005.
71. C. Si, W. Chen, W. Wang, L. Wang, and T. Tan, "An attention enhanced graph convolutional lstm network for skeleton-based action recognition" , 2019, in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2019, pp. 1227–1236.
72. A. B. Sargano, X. Wang, P. Angelov, and Z. Habib, "Human action recognition using transfer learning with deep representations" , 2017, in *International joint conference on neural networks (IJCNN)*, 2017, pp. 463–469.
73. A. Shahroudy, T.-T. Ng, Y. Gong, and G. Wang, "Deep multimodal feature analysis for action recognition in rgb+ d videos", 2017, *IEEE transactions on pattern analysis and machine intelligence*, vol. 40, no. 5, pp. 1045–1058.
74. D. Avola, M. Bernardi, and G. L. Foresti, "Fusing depth and colour information for human action recognition" , 2019, in *Multimedia Tools and Applications*, vol. 78, no. 5, pp. 5919–5939.
75. H. Naveed, G. Khan, A. U. Khan, A. Siddiqi, and M. U. G. Khan, "Human activity recognition using mixture of heterogeneous features and sequential minimal optimization" , 2019, in *International Journal of Machine Learning and Cybernetics*, vol. 10, no. 9, pp. 2329–2340

AUTHORS PROFILE

Mallikarjun Aralimarad obtained his B.E degree in Electronics and communication Engineering from HIT, Nidasoshi, affiliated to VTU, Belagavi, Karnataka and obtained his M.Tech in the area of VLSI design and Embedded Systems from BVBCET,Hubli, affiliated to VTU Belagavi, Karnataka. Currently he is pursuing Ph.D in the area of Image and Video Processing. His Areas of interest are Image and Video Processing, VLSI. He is Assistant professor of Electronics and Communication Engineering, Basaveshwara Engineering College (Autonomous), Bagalkot, Karnataka, India.

.

Dr. Meena S. M. obtained her B.E (Electronics and Communications engineering) degree from Karnataka University Dharwad, M.Tech(Communications Systems) from IIT, Roorkee in 1995 and PhD (Information systems) from VTU, Belagvi, in 2012.She has published more than 35 publications. Her areas of interest include information retrieval, Data analytics. Presently, she is a HOD and Professor of School of Computer Science and Engineering in KLETECH, Hubli.

.

Dr. Jayashree D. Mallapur obtained her B.E (Electronics and Communication Engineering) degree from Karnataka University Dharwad in 1991 and M.Tech from Poojya Doddappa Appa College of Engineering, Kalaburagi, (Gulbarga University, Gulbarga) in 1995 & Ph.D (Fuzzy based resource allocation and multicast routing in wireless multimedia cellular networks) from Gogate Institute of Technology, Belagavi, (VTU, Belagavi) 2009.She has published more than 51 publications. She is a Professor of Electrictronics and Communication Engineering department at Basaveshwar Engineering College (Autonomous), Bagalkot.