IOSR Journal of Computer Engineering (IOSR-JCE)
e-ISSN: 2278-0661,p-ISSN: 2278-8727, Volume 19, Issue 4, Ver. VI. (Jul.-Aug. 2017), PP 43-50
www.iosrjournals.org

A Survey on Human Action Recognition

Ayush Purohit [*], Shardul Singh Chauhan[*]
[#]*Centre for Information Technology, University of Petroleum and Energy Studies Dehradun, India*
Corresponding Author: Ayush Purohit

Abstract: *Human Action Recognition is one of the important research area in computer vision and image processing field. A machine can be interacted and controlled by gestures and facial expression using the visual modality. Human action recognition can be seen as a bridge for machines to understand human body language, an intelligent interacting system between machines and humans which limits majority of input devices such as keyboard and mouse. Thus, producing an accurate and meaningful information on human activities and behaviours is one of the important tasks in pervasive computing. In this survey, we tried to focus on advancement of action recognition after yr. 2010.*

Keywords: *Human Action recognition, Gesture Recognition, Human Computer Interaction*.

Date of Submission: 04-08-2017 Date of acceptance: 17-08-2017

I. Introduction

To live in modern society, intelligent frameworks and administrations are required to support our collective and informative needs as social beings. Development and Improvements in programming and equipment innovations, for example, in microelectronics, mechatronics, discourse innovation, semantics, computer vision, and computerized reasoning are consistently driving new applications for work, relaxation, and versatility [1]. In computer vision, human action recognition is one of the most popular research area. With the technology melioration, it is required to forge an entity which can process the prodigious volume of visual records and prosecute auto-scrutiny. Some of its application comprises in vigilance systems, video analysis, web-video search and retrieval, quality-of-life devices for elderly people and a variety of systems that involves man-machine interaction [1] [2].

Human action can be defined as any specific behaviour asserted by the human body. In the prevailing gimmick, recognition of human behavior is imperative, but arduous job. Action recognition is imperative by seeing it as a resilient and visceral approach to promote more human-centred forms of man-machine interaction. However, the effort required to function these recognitions varies and are very complex due to wide range of sub goals such as unique identification of body parts, recognising gestures and classifying them. A basic methodology to recognise a human action is to detect human, segmenting and mapping the body attributes followed by the results.

There is literature, previously done on human action recognition which involves different approaches to measure human actions. In [4], a review on full body actions is done. In this, the classification of action is studied on the basis of spatial and temporal structure of body movements. In [2] [3], hierarchal approach is reviewed to differentiate the human action recognition problems. In the review, human action is classified into two ways i.e., single layered approach and hierarchal approach. Based on the different approaches, there are further classification to recognize human actions which are based on different representation and learning methodologies as shown in figure 1.

Volker Kruger [5], categorised and reviewed on the basis of scene, full body and with/without using body parts. Jose M. Chaquet and Enrique J. Carmona reviewed 68 datasets available for human action recognition within the time interval of 2001 to 2012. In this paper, it was concluded that out of all dataset available, Weizmann, KTH and CAVIAR are the most popular dataset used worldwide for research and development purposes. In [6], usage of fusing depth cameras and internal sensors is outlined which provides 3 dimensional structure of human actions. Additionally a review of publically available dataset has been done. Teofilo E. de Campos[7] survey focus on the applications of the human action recognition in video survilliance for observing people in real time. An approach for attribute-based person recognition from large scale monitoring structure is also included in the same survey.

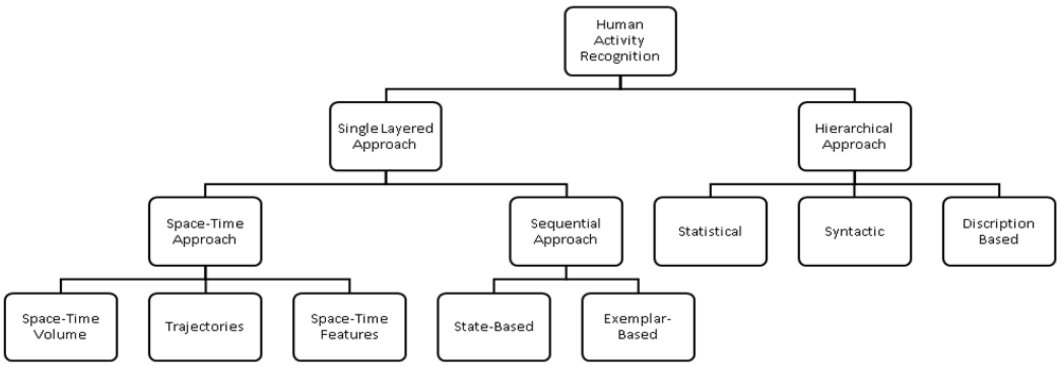

Figure 1: *Classification of human activity recognition [2].*

This paper thrust on new researches ideologies which are not discussed in previous surveys. In this survey, Human action recognition are categorised into three parts [8] i.e., Human body model based method, Holistic methods, local feature method. In section II, Human body based model is discussed thereafter Holistic methods for recognizing the human actions. Local feature based methods are then discussed in section IV. At last we concluded this survey in section V.

II. Human Body Model Based Method

The human body model based methodology of action recognition makes utilization of the bits of data which is extracted from the body parts of the human beings. The model mainly comprises of the two necessary principles [42]: The appearance of the body part of the human being in the image or video [43]. With the assistance of moving light display, the people can perceive the motions by depicting the movements of the fundamental joints of individuals. In the study performed by Johansson [42], established that the range between ten to twelve of the moving light displays which are in movement mixes in proximal boost have evoked the impression of human body motions. An extended approach of epipolar geometry that is the geometry of dynamic scenes is indicated in [43]. The system has been implemented in variety of sequences. The theory of chaotic systems in [44] have been used for action recognition. A complete set of new features are implemented in order to classify human body gestures that are dynamic in nature.

A moving pose framework, consist of moving pose descriptor (MP) and a non-parametric kNN classifier, is proposed in [9] for recognizing actions in unsegmented order. To capture and represent 3D kinematic pose, frame descriptors are used along with MP descriptors, sequenced in time. To measure the differential quantity of joints, velocity of human pose is estimated thereafter pose normalization is performed to reduce noise from video frames. Classification of actions is performed during training to remove unrecognised actions from the video frame using kNN. In [10], proposed Bag of Poses classification model with weak poses for recognising human actions. To recognize 3D actions from a frame, action specific motion modelling is done thereafter PCA is applied to all the actions in order to remove the redundancy between the poselets. To compute a weak pose, SCD is used to place a sampled point on a shape in the origin of a radial coordinate system which are encoded into 5 bins of histogram. The calibration between rigid actions and weak poses is done via GPR and energy k-means is used for computing action's vocabulary. For classifying actions, SVMs are used which is tested upon HumanEva and IXMAS dataset and produces nearly 79.2% action recognition rate.

Jamie et al. [11], Introduced an approach to vaticinate 3D positions of body joints from a single depth static image. A highly diverse dataset of mocap allows deep decision forests to classify the human actions using depth invariable lineaments without overfitting. Using mean-shift with a weighted Gaussian kernel, per-pixel information is pooled across the pixels to produce 3D skeletal joints over the surface of body. The overall performance of the system results in achieving 0.731 mAP in synthetic dataset while in real test dataset 0.984 mAP is achieved when ground truth of the body parts is known. Lu Xia, Chia-Chih and Aggarwal in [12] presented an invariable portrayal of actions using HOJ 3D descriptor. A histogram based representation of actions in which a posture is calculated using 12 joints. Using Kinect, these skeletal joint locations are extracted after which LDA is implemented to extracted dominant features. To classify sequential postures into actions, discrete HMM is trained, which results in overall 96.2% of recognition rate when 1/3rd of the MSRAction3D dataset samples are taken as training sets. The system produces 78.97% of accuracy in a cross subject test when half of these samples are used as training and rest are taken as testing sets.

A real-time tracking based human action recognition is proposed in [13]. In this system depth frames are acquired using Kinect sensors in real time, which are then processed using OpenNI skeleton tracking algorithm. For pose estimation, orthogonal pose vectors are evaluated for every frame which is thought-out to be the key feature to evaluate any human posture. Angle of joints are used to represent posture which is used

with motion energy-based approach to achieve horizontal symmetry. To classify the processed result, HMM along with GMM is used to recognize human actions. A mid-level representation [14] is proposed alluded as dynamic postures. The dynamic stances are fit for coupling the neighbour data of movement autonomously and straightforwardly with the skeletal stances. A separation capacity for the represents that are alterable have been recommended. The prescribed technique which have been recommended makes utilization of element stances over crude skeletal stances in movement recognition utilizing bolster vector machine and codebook coordinating classifiers.

The activities are arranged based upon the stances, movement and appearance given that the information of the video or time arrangement of skeleton has been given in [15]. The utilization of the structure has been to take in the movements which are skeletal and discriminative movement designs with the end goal of activity affirmation. A body part movement based element has been suggested likewise alluded as moving poselets have been utilized. Moving poselets compares to the body part setup while experiencing a particular development. A calculation has been proposed for the learning procedure of moving poselets and activity classifiers. The last result gives an unmistakable sign that the proposed technique offers a high recognition rate when contrasted with other existing methodologies.

III. Holistic Based Method

The holistic based technique for action recognition uses the data retrieved on individual's localization in real time datasets [45]. The silhouette and in addition the shape based components serves as one of the principle characteristic that have been utilized to represent the human dynamics and the body structure for recognition of actions in videos [51] [52]. The fundamental point of the holistic based strategy is not to utilize any of the data provided by the body parts of the people.

An appearance and motion cues based human action recognition is implemented in [16]. The proposed methodology initialised by recognizing region of interest using EdgeBox on optical flow images. Every EdgeBox on the image gets a grade based on normalised magnitude of the optical flow signal which thenceforth used as input to the nets. A VGG-16 layered architecture based on convolutional model is selected for training spatio-temporal features separately. CNN features extracted are then fused and trained using linear SVM. A semiholistic approach for key pose selection and representation is proposed in [17] for recognizing human actions. For every video frame, EPFs are extracted to evaluate orientation, intensity and contour using image processing techniques. A subset of poses is selected to abolish redundancy which is adept using supervised AdaBoost learning algorithm. Weighted LNBNN classifier is proposed which employs kNN search, the algorithm curtail overall time to classify human actions and outperform state-of-art techniques with an average recognition rate of 94.8% in KTH dataset.

In [18], two procedures have been proposed to extract the relevant features for recognizing actions. This procedure involves all-encompassing based elements to be extricated which are applied on a few information sets and extricating the focuses where the joints of individuals are found in poselets. An effectively accessible information MPI dataset has been utilized for recognition of activities. A correlation is additionally done so as to figure out the best system. A holistic based all-encompassing methodology is proposed in [19]. To represent human poselets, movements are consolidated in pixel change history image which is accomplished by part the pixel change history into different channels that are arranged comparing to distinctive directions. Invariant features are extricated from pixel change history images and Naive Bayesian model has been utilized for the procedure of recognition.

Another method has been suggested in [20] for activity recognition based on all-encompassing elements and neighbourhood descriptors. The activity recognition structure utilizes outline differencing, highlight combination and sack of words methods. The two sorts of nearby descriptors i.e., 2D and 3D scale invariant feature transform for depiction of elements are utilized. Finally, SVMs has been utilized as a part of the work to train and test the framework. Xinxiao Wu et. al. [21] proposes spatio temporal context distribution feature for recognizing human activity in a temporal framework. To identify the dispersion of spatio-temporal feature set in video frames, multiple GMMs are implemented from different space-time scales obtained from global GMM using MAP. To train the model, AFMKL learning is proposed for coalescing spatio-temporal and local appearance dissemination characteristics which can adapt to multiple kernels by training it against different features.

A human activity recognition in [22] has been developed by utilizing inertial sensor alongside profundity camera. Both the depth image and inertial sign elements are effective in terms of processing and recognition rate. These elements are actualized in computationally effective two collective classifiers on which a decision level combination must be performed. An activity signature based methodology in [23] has been proposed for recognizing human activities. A 1D direction is analysed by parsing 2D picture which contain movement highlights outline alluded as movement history picture. This direction signifies the change of the movement as for time. A system based upon the mixture of dynamic programming and Von Mises conveyance

for arrangement of successions are utilized to characterize the directions. A large portion of the work done in the field of activity recognition has been on the successions which are procured by the cameras which are stationary with some settled measure of interest focuses [49].

At the point when the camera moves the directions took after by the different parts of the body contains not just the movement of the performing artist in the watched scene additionally the camera in movement [50]. In expansion to the movement of the camera the interest purposes of some activity in various situations results in an alternate number of directions. With a specific end goal to handle such sort of circumstances a multi view geometry between the two activities have been proposed in [24].

A technique in [25], the system which has been utilized for recognition of activities of people where the elements must be separated from the districts encompassed alluded as an adverse space. The framework utilizes a division plan which is various levelled in nature, disposal of shadows, highlight extraction based upon shape, figuring of velocity, parcels of the areas and coordinating of grouping by time distorting which is alert. A similar approach is used in [26] [27], recognition of activities is performed in a grouping of video of discretionary length. Based upon the representation of sack of words and in addition idle semantic investigation demonstrate the procedure continues by edge and the redesigns of choice every once in a while.

IV. Local Feature Based Method

The local feature based approach implements the usage of local features for the process of action recognition. The most important principle involved is that there is no need of the data on localization of people or data of the model of human body. This technique has been a standout amongst the most exceptionally examined area in the field of action recognition.

A two-stream deep CNN architecture combined using spatial stream CNN and temporal stream CNN is proposed by Karen and Andrew [28]. The spatial CNN hoist information about the objects described in a single video frame whereas the temporal CNN carries the movement of observer and objects across the frames. Optical flow displacement field is used to represent actions. To classify the actions, softmax scores is fused using either averaging or linear SVM to minimize over-fitting. Using SVM as a fusion method in unidirectional, multitasking temporal CNN, the recognition rate for this architecture is found to be maximum i.e., 87%. Bag of features in [47] have been utilized for recognizing human actions. The functionality of bag of features is in projecting a particular feature to a local co-ordinate system. Further the co-ordinates that are projected are integrated by using max pool approach to produce final system. Fisher vector encoding method in [48] makes use of the differences between the visual words and the features. Initially it creates a visual vocabulary where process of clustering is utilized. Later on, differences that are of second and first order between the visual vocabulary as well as local features.

In [29], developed a 3D- CNN model which extracts features form both spatial and temporal dimensions. This model captures motion which then performs convolution and subsampling in different channel using contiguous frames. All channels are then calculated by merging to produce feature representation of the action. TRECVID dataset used for training, a bounding box for every human action within the video frame was calculated and extracted by the 3D CNN, leading to a cube which describes action of a single person. BoW feature is then calculated for each cube and classification of actions is done using linear SVM. Du Tran et al. in [30], proposed 3D CNN for spatial feature learning in a large scale supervised video dataset. To combine temporal information using deep nets, homogeneous temporal depth and varying temporal depth architectures are developed to study the behaviour of temporal depth in CNN. A 3D CNN with 3*3*3 kernel is used for spatiotemporal feature leaning. After training the architecture using SGD, C3D video descriptor is used to extract features. Linear multi-class SVMs are used for training and classifying action features. Recognition rate of human actions are measured Upto 90.4% which uses combination of C3D (3 nets), iDT and linear SVM.

Pyry, Rahul and Sukthankar in [31], introduced feature seeding technique, which uses synthetic data to improve the rate of action recognition. In this, synthetic data is used for selecting latent features over the real-world dataset. To do that, a threshold should be met for deciding the similarity between similar types of features which can be used in synthetic and real-world dataset. This technique produces overall maximum recognition rate to 68% in MSR dataset. Karinne et al. in [32], presented a technique to recognize human actions using extended ISA algorithm with PCA, which elicit spatio temporal features from 3D video blocks. The SCN model is trained greedily layer-wise over the network thereafter classification of local features is performed using SVMs. In order to infer activities, human actions and information of objects are observed. This information is then used to model a decision tree using C4.5 algorithm to classify the activities.

Earnest and Krishna in [33], proposed fuzzy CNN for identifying human actions in temporal domain. To overcome the issue of alignment and complex limb operations for recognizing an action, feature extraction and classification of action components are evaluated using fuzzy membership functions and MOCAP information. Three joints (right/left hand & pelvis) are considered for tracking this information to compute displacement between right-left hand and between the joints to the ground which is then normalised using its

reference pose. The utilization of layered representation for the procedure of action recognition has been suggested in [34]. The fundamental point has been to execute a layered based probabilistic representation utilizing concealed markov model for performing the assignment of detecting, inferencing and to learn at numerous levels of worldly granularity. The utilization of representation in the prescribed framework for the conclusion of conditions of the action of the client based upon the constant surges of video, acoustic and PC connection.

The assignment of human activity perception includes the predominance of the mid-level and in addition low level elements. Focus on action realization in video utilizing abnormal state highlights has been named as Action Bank [35]. The activity bank includes on a hefty portion of the considerable activity locators which can be extensively tested in the perspective and semantic space. The proposed representation has been matched with the bolster vector machine which is straight in nature whose execution has been great. A procedure of executing a spotting and perceiving constant human action realization framework by utilizing a vision sensor has been recommended in [36]. The recommended technique contains profundity MHI HOG, spotting of activities, displaying of activities and recognition. The principal errand to dispense with the closer view from the foundation of the picture has been by utilizing profundity MHI HOG. The second errand demonstrates the various types of activities by utilizing the removed components. The creating so as to display of activities executes the arrangements of activities with the assistance of k means clustering. These arrangements have been utilized as an information to the shrouded Markov model.

The significant issue for recognition of human activities emerges when the body movements changes because of variety in degrees of movement of people. The part of cooperative energies inside of the model of activity recognition is focused in [37]. The subject of how collaboration can be utilized for the procedure of activity recognition have likewise been replied in the proposed work. The utilization of cooperative energies inside of the generative construction modelling of activity execution and additionally recognition has been one of the prescribed work. A bed adjusted guide descriptor has been discussed in [38]. The descriptor has been security cognizant, free of alignment representation of a solitary individual bed which has been gotten from a profundity camera. This serves an exceptionally handy method for observing uninterruptedly. The methodology has been utilized for recognition of both time variation and also static. The work depicts the abnormal state representation of the bed adjusted guide descriptor for an abnormal state of comprehension of the scene.

Deep convolutional systems have discovered its way to the examination done in the field of grouping of images and in addition movement affirmation, one such work is described in [39] [46] [53]. For the fair-minded mechanical autonomy applications, the degree of variety and advancement in real life foundations has been far higher than the information sets of PC vision. One such framework proposed in [36] which removes the locales from the picture which are expectedly fit for discovering the activities both in the periods of testing and preparing. The undertaking of perceiving activities in a video has been focused in [40]. The introductory work done in this field are based upon the measurements of the nearby video highlights. The work highlights the representation significance which has been gotten from human stances. One of the alternative procedure in particular the stance based recurrent neural network descriptor has been recommended for recognition of activities [54]. Different methodologies of transient accumulation and the stance based recurrent neural networks highlights have been gotten from the physically explained and consequently assessed human stances.
Recurrent neural networks serve as a proficient strategy for anticipating and ordering the groupings. Recurrent neural networks can be utilized successfully as a part of request to encode the groupings and to give a powerful representation [41]. The approach utilized has been based upon the fisher vectors. In the fisher vectors the recurrent neural networks usefulness are on the era of the probabilistic models and the calculation of the halfway subsidiaries has been finished by utilizing back spread. The trial investigation has been finished by utilizing the successions. The arrangements utilized here are the picture annotation and video activity recognition.

V. Discussion And Conclusion

In this paper, we presented an extensive survey by summarizing the explored research techniques in the field of human action recognition. The survey reveals the milestones achieved during the past five or six years in recognising human actions and activities. We classified the existing techniques based on information on human body parts, people localization and local features. However, there still remain various arduous issues for perfecting the system and deploying in real-world. Some of the limitations which are common among various approaches includes camera calibrations, natural gestures, moving human segmentation and action vocabulary. These problems mainly arise due to insufficient realistic datasets and restricted environments while developing the system.

TABLE *1* HUMAN RECOGNITION CHART BETWEEN DIFFERENT APPROACHES

Method	Author	Key Features	MSR Action 3D Dataset	KTH Dataset	Others
Model Based Approach	Mihai Zanfir et. al. [9]	Non-Parametric Moving Pose Framework.	91.70%	-	MSRDailyActivity: 73.80%
	Wenjuan Gong et. al. [10]	Weak Pose Estimation.	-	-	HumanEva: 91.1% IMAX: 79.20%
	Jamie Shotton et. al. [11]	Body part inferencing using random decision forests.	-	-	Synthetic Dataset: 0.731 mAP.
	Lu Xia et. al. [12]	View invariant 3D skeletal joint representation.	96.20%	-	-
	Georgios Th. Papadopoulos et. al. [13]	Action representation based on spherical angles between joints.	-	-	Grand Challenge: 76.03%
	Ran Xu et. al. [14]	Dynamic Pose representation.	-	91.2%	UCF-Sports: 81.33%
	Lingling Tao et.al. [15]	Learning framework for discriminative and interpretable patterns.	93.60%	-	MSRDailyActivity: 74.50% Berkeley MHAD: 100%
	A Yilma et. al. [43]	Multi-view geometry between 2 actions.	-	-	-
	Saad Ali et. al. [44]	Analysing non-linear dynamics using chaotic model.	-	-	Weizmann: 92.6%
Holistic Approach	Li Liu et. al. [17]	Representing human posture using Extensive Pyramidal Features.	-	94.8%	Weizmann: 100% IMAX: 94.5% HMDB51: 49.3%
	Xinghua Sun et. al. [20]	Unified action recognition framework fusing local descriptors and holistic features.	-	94.0%	Weizmann: 97.8%
	Xinxiao Wu et. al. [21]	Context distribution feature based on STIPs.	-	94.5%	UCF Sport: 91.3%
	Chen C. et. al. [22]	Use of inertial sensor and depth camera on previously developed sensor fusion method.	-	-	UTD-MHAD: >97%
	Calderara S. et. al. [23]	Novel holistic representation of actions based upon motion history image.	-	-	Subset of frames collected from 25 videos: >95%
	Yilma A. et. al. [24]	Use of geometry of dynamic scenes for the cameras in motion	-	-	Own database: Good accuracy rate
	Shah Atiqur Rahman et. al. [25]	Region based technique to recognize human actions in negative space.	-	95.49%	Weizmann: 100%
	Yu-Gang Jiang et. al. [26]	Bag of features representation.	-	-	Hollywood2- 59.5% Olympic Sports-80.6% HMDB51- 40.7%
	Ping Guo et. al. [27]	Use of bag of words representation as well as probabilistic latent semantic analysis model.	-	91.8%	Weizmann: 81.3%
	J. Yamato et. al. [45]	Feature based bottom up approach.	-	-	10 mixed data sets-78.5 % and 70.8%
	Mona M. Moussa et. al. [49]	Detection of interest points using SIFT technique.	-	97.89%	Weizmann: 96.66%
	Ning Xu et. al. [50]	Accuracy under multi view scenario.	-	-	M^2I:75.7%/72.8% and 76.5%/ 75.4%
	Di Wu et. al. [51]	Correlation between actions.	-	-	IMAX:83.6%, 90.3%, 89.4%, 89.8% and 78.8%
Local Feature Based	Karen Simonyan et. al. [28]	Use of convolution neural network.	-	-	UCF-101: 87.0% HMDB-51: 81.5%
	Shuiwang Ji et. al. [29]	Extraction of both temporal and spatial dimensions.	-	-	TRECVID Data: 90.2%
	Du Tran et. al. [30]	Spatio temporal feature learning.	-	-	UCF-101: 52.8%
	Pyry Matikainen et. al. [31]	Synthetic data to search for robust features.	-	-	MSR: 68.0%
	Karinne Ramirez-	Automatic generation of	-	-	Own dataset: 82%

	Amaro et. al. [32]	semantic rules.			
	Earnest Paul Ijjina et. al. [33]	MOCAP Tracking.	-	-	MHAD: 99.248%
	Sreemanananth Sadanand et. al. [34]	Action Bank	-	98.2%	UCF Sports: 95.0% UCF50: 57.9% HMDB-51: 26.9%
	Fahimeh Rezazadegan et. al. [35]	Action region informed by optical flow.	-	-	UCF101: 69.98%
	Nuria Oliver et. al. [36]	Use of layered probabilistic representations.	-	-	Own dataset:72.68%, 99.7%
	Hyukmin Eum et. al. [37]	Depth MHI HOG.	-	-	Weizmann: 98.21%
	Manuel Martinez et. al. [38]	Use of BAMs.	-	-	BAM Dataset: 97%
	Guy Lev et. al. [40]	Use of fisher vectors I RNNs.	-	-	HMDB-51: 67.71% UCF101: 94.08%
	Guilhem Cheron et. al. [41]	Aggregation of motion and appearance information of parts of human body.	-	-	JHMDB: 79.5%, 72. 2% MPII Cooking: 71.4%

References

[1] Karl-Friedrich Kraiss, Advanced Man-Machine Interaction Fundamental & Implementation, © Springer-Verlag Berlin Heidelberg, 2006.

[2] Aggarwal, J., Ryoo, M., Human activity analysis: A survey, ACM Computing Surveys 43, 1-43, 2011.

[3] Guangchun Cheng, Yiwen Wan, Abdullah N. Saudagar, Kamesh, Advances in Human Action Recognition: A Survey, arXiv: 1501.05964v1 [cs.CV] 23 Jan 2015.

[4] Daniel Weinlanda, Remi Ronfardb, Edmond Boyerc, A Survey of Vision-Based Methods for Action Representation, Segmentation and Recognition, volume 115, Issue 2, February 2011, Pages 224-241.

[5] Volker Kruger, Danica Kragic, Ales Ude, Christopher Geib, The Meaning of Action: A review on action recognition and mapping, Taylor and Francis Online, volume 21, Issue 13, 2007.

[6] Chen Chen, Roozbeh Jafari, Nasser Kehtarnavaz, A survey of depth and inertial sensor fusion for human action recognition, © Springer Science + Business Media New York, 2015.

[7] Teofilo E. de Campos, A survey on computer vision tools for action recognition, crowd surveillance and suspect retrieval, XXXIV Congresso da Sociedade Brasileira de Computação – CSBC 2014.

[8] Piotr Tadeusz Bilinski, Human action recognition in videos, Computer Vision and Pattern Recognition [cs.CV], Universite Nice Sophia Antipolis, 2014.

[9] Mihai Zanfir, Marius Leordeanu, Cristian Sminchisescu, The Moving Pose: An Efficient 3D Kinematics Descriptor for Low-Latency Action Recognition and Detection, IEEE Conference on Computer Vision, 2013.

[10] Wenjuan Gong, Jordi Gonzàlez and Francesc Xavier Roca, Human action recognition based on estimated weak poses, EURASIP Journal on Advances in Signal Processing, 2012.

[11] Jamie Shotton, Andrew Fitzgibbon, Mat Cook, Toby Sharp, Mark Finocchio, Richard Moore, Alex Kipman, Andrew Blake Real-Time Human Pose Recognition in Parts from Single Depth Images, IEEE Conference on Computer Vision and Pattern Recognition, 2011.

[12] Lu Xia, Chia-Chih Chen, and J. K. Aggarwal, View Invariant Human Action Recognition Using Histograms of 3D Joints, IEEE Computer Society Conference on Computer Vision and Pattern Recognition Workshops (CVPRW), 2012.

[13] @Inbook Georgios Th. Papadopoulos, Apostolos Axenopoulos and Petros Daras, chapter Real-time Skeleton-tracking-based Human Action Recognition Using Kinect Data, MultiMedia Modeling: 20th Anniversary International Conference, MMM 2014, Dublin, Ireland, January 2014, Proceedings Part I, pages 473—483, Springer International Publishing, 2014.

[14] Ran Xu, Priyanshu Agarwal, Suren Kumar, Venkat N. Krovi, Jason J. Corso, Combining Skeletal Pose with Local Motion for Human Activity Recognition, 7th International Conference, Springer, 2012.

[15] Lingling Tao, Rene Vidal, Moving Poselets: A Discriminative and Interpretable Skeletal Motion Representation for Action Recognition, IEEE Conference on Computer Vision Workshops, 2015.

[16] Fahimeh Rezazadegan, Sareh Shirazi, Niko Sünderhauf, Michael Milford, Ben Upcroft, Enhancing Human Action Recognition with Region Proposals, Journal of Advanced Research, March 2015.

[17] Li Liu, Ling Shao, Xiantong Zhen, and Xuelong Li Learning Discriminative Key Poses for Action Recognition, IEEE Transactions on Cybernetics, Jan 2013.

[18] Leonid Pishchulin, Mykhaylo Andriluka, Bernt Schiele, Fine-grained Activity Recognition with Holistic and Pose based Features, Max Planck Institute for Informatics, Germany, Stanford University, USA.

[19] Jia-xin Cai, Guo-can Feng, Xin Tang, Human action recognition using oriented holistic feature, IEEE International Conference on Image Processing (ICIP), 2013.

[20] Xinghua Sun, Mingyu Chen, Hauptmann A., Action recognition via local descriptors and holistic features, IEEE Computer Society Conference on Computer Vision and Pattern Recognition Workshops, 2009.

[21] Xinxiao Wu, Dong Xu, Lixin Duan, Jiebo Luo, Action recognition using context and appearance distribution features, IEEE Conference on Computer Vision and Pattern Recognition, 2011.

[22] Chen C., Jafari R., Kehtarnavaz N., A Real-Time Human Action Recognition System Using Depth and Inertial Sensor Fusion, IEEE Sensors Journal, October 2015.

[23] Calderara S., Cucchiara R., Prati A., Action Signature: A Novel Holistic Representation for Action Recognition, IEEE Fifth International Conference on Advanced Video and Signal Based Surveillance, 2008.

[24] Yilma A., Shah M., Recognizing human actions in videos acquired by uncalibrated moving cameras, Tenth IEEE International Conference on Computer Vision, 2005.

[25] Shah Atiqur Rahman, M. K. H. Leung, Siu-Yeung Cho, Human Action Recognition by Extracting Features from Negative Space, Springer Berlin Heidelberg, 2011.

[26] Yu-Gang Jiang, Qi Dai, Xiangyang Xue, Wei Liu, Chong-Wah Ngo, Trajectory-Based Modeling of Human Actions with Motion Reference Points, Springer Berlin Heidelberg, 2012.

[27] Ping Guo, Zhenjiang Miao, Yuan Shen, Heng-Da Cheng, Real Time Human Action Recognition in a Long Video Sequence, Seventh IEEE International Conference on Advanced Video and Signal Based Surveillance, 2010.

[28] Karen Simonyan, Andrew Zisserman, Two-Stream Convolutional Networks for Action Recognition in Videos, NIPS, 2014. arXiv:1406.2199v2.

[29] Shuiwang Ji, Wei Xu, Ming Yang, Kai Yu, 3D Convolutional Neural Networks for Human Action Recognition, IEEE Transactions on Pattern Analysis and Machine Intelligence, March 2012.

[30] Du Tran, Lubomir Bourdev, Rob Fergus, Lorenzo Torresani, Manohar Paluri, Learning Spatiotemporal Features with 3D Convolutional Networks, arXiv:1412.0767v4 [cs.CV], Oct 2015.

[31] Pyry Matikainen Rahul Sukthankar Martial Hebert, Feature Seeding for Action Recognition, International Conference on Computer Vision, 2011.

[32] Karinne Ramirez-Amaro, Eun-Sol Kim, Jiseob Kim, Byoung-Tak Zhang, Michael Beetz and Gordon Cheng, Enhancing Human Action Recognition through Spatio-temporal Feature Learning and Semantic Rules, 13th IEEE-RAS International Conference on Humanoid Robots (Humanoids), 2013.

[33] Earnest Paul Ijjina, C Krishna Mohan, Human Action Recognition based on Motion Capture Information using Fuzzy Convolution Neural Networks, Eighth International Conference on Advances in Pattern Recognition (ICAPR), 2015.

[34] Sreemanananth Sadanand, Jason J. Corso, Action Bank: A High-Level Representation of Activity in Video, IEEE Computer Society Conference on Computer Vision and Pattern Recognition, 2012.

[35] Fahimeh Rezazadegan, Sareh Shirazi, Niko Sünderhauf, Michael Milford, Ben Upcroft, Enhancing Human Action Recognition with Region Proposals, Proc. of Australasian Conference on Robotics and Automation (ACRA), 2015.

[36] Nuria Oliver, Eric Horvitz, Ashutosh Garg, Layered Representations for Human Activity Recognition, Fourth IEEE International Conference on Multimodal Interfaces, 2002.

[37] Hyukmin Eum, Changyong Yoon, Heejin Lee, Mignon Park, Continuous Human Action Recognition Using Depth-MHI-HOG and a Spotter Model, Sensors, 2015.

[38] Manuel Martinez, Lukas Rybok, Rainer Stiefelhagen, Action Recognition in Bed using BAMs for Assisted Living and Elderly Care, 14th IAPR International Conference on Machine Vision Applications, 2015.

[39] Pezzulo G, Donnarumma F, Iodice P, Prevete R, Dindo H, The role of synergies within generative models of action execution and recognition: A computational perspective, Physics of Life Reviews, 2015.

[40] Guy Lev, Gil Sadeh, Benjamin Klein, Lior Wolf, RNN Fisher Vectors for Action Recognition and Image Annotation, CoRR, 2015.

[41] Guilhem Cheron, Ivan Laptev, Cordelia Schmid, P-CNN: Pose-based CNN Features for Action Recognition, IEEE Conference on Computer Vision Workshops, 2015.

[42] Gunnar Johansson, Visual perception of biological motion and a model for its analysis, Perception & Psychophysics, Springer, vol. 14, no. 2, pages 201–211, 1973.

[43] A Yilma and Mubarak Shah, Recognizing human actions in videos acquired by uncalibrated moving cameras, Tenth International Conference on Computer Vision, volume 1, pages 150–157, IEEE, 2005.

[44] Saad Ali, Arslan Basharat, Mubarak Shah, Chaotic invariants for human action recognition, 11th International Conference on Computer Vision, pages 1–8, IEEE, 2007.

[45] J. Yamato, J. Ohya, K. Ishii, Recognizing human action in time sequential images using hidden Markov model, Computer Society Conference on Computer Vision and Pattern Recognition, pages 379–385, IEEE, 1992.

[46] Karen Simonyan, Andrew Zisserman, Very Deep Convolutional Networks for Large-Scale Image Recognition, arXiv:1409.1556v6.

[47] Jinjun Wang, Jianchao Yang, Kai Yu, Fengjun Lv, Thomas S. Huang, Yihong Gong, Locality-constrained Linear Coding for image classification, IEEE Conference on Computer Vision and Pattern Recognition, pages 3360–3367, IEEE, 2010.

[48] Florent Perronnin, Jorge Sánchez, Thomas Mensink. Improving the Fisher Kernel for Large-Scale Image Classification Proceedings of the 11th European conference on Computer vision, 2010.

[49] Mona M. Moussa, Elsayed Hamayed, Magda B. Fayek, Heba A. El Nemr, An enhanced method for human action recognition, Journal of Advanced Research, Volume 6, Issue 2, Pages 163–169, March 2015.

[50] Ning Xu, Anan Liu, Weizhi Nie, Yongkang Wong, Fuwu Li, Yuting Su, A Multi-modal & Multi-view & Interactive Benchmark Dataset for Human Action Recognition, Proceedings of the 23rd ACM international conference on Multimedia, Pages 1195-1198, 2015.

[51] Di Wu, Ling Shao, Silhouette Analysis-Based Action Recognition via Exploiting Human Poses, IEEE Transactions on Circuits and Systems for Video Technology, Volume 23, Issue 2, 2013.

[52] Mintao Zhao, William G. Hayward, Isabelle Bülthoff, Holistic processing, contact, and the other-race effect in face recognition Vision Research, Volume 105, December 2014, Pages 61–69.

[53] Alex Krizhevsky, Ilya Sutskever, Geoffrey E. Hinton, ImageNet Classification with Deep Convolutional Neural Networks, Advances in Neural Information Processing Systems 25 (NIPS 2012).

[54] Jeffrey Donahue, Lisa Anne Hendricks, Sergio Guadarrama, Marcus Rohrbach, Subhashini Venugopalan, Kate Saenko, Trevor Darrell; The IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2015, pp. 2625-2634.

IOSR Journal of Computer Engineering (IOSR-JCE) is UGC approved Journal with Sl. No. 5019, Journal no. 49102.

Ayush Purohit . "A Survey on Human Action Recognition." IOSR Journal of Computer Engineering (IOSR-JCE), vol. 19, no. 4, 2017, pp. 43-50.