© 2018 IJSRST | Volume 4 | Issue 5 | Print ISSN: 2395-6011 | Online ISSN: 2395-602X
Themed Section: Science and Technology

# Human Activity Recognition and Classification – A Comprehensive Survey

**G. P. Hegde, Nireeksha, Nithini B. K, Soundarya M. B, Supreetha B. V.**
Department of Information Science, SDMIT, Ujire, Karnataka, India

## ABSTRACT

Human action recognition from realistic videos attracts more attention in many practical applications such as on-line video surveillance and content-based video management. Single action recognition always fails to distinguish similar action categories due to the complex background settings in realistic videos. This paper explains the various studies on human activity recognition and also focuses on steps needed for recognition of human activity using local invariant methods and classification of human activity frames.

**Keywords:** Activity recognition, Local space, classifier

## I. INTRODUCTION

EApplications such as surveillance, video retrieval and human-computer interaction require methods for recognizing human actions in various scenarios. Typical scenarios include scenes with cluttered, moving backgrounds, non-stationary camera, scale variations, individual variations in appearance and cloth of people, changes in light and view point. This work demonstrates how the action recognition can be achieved using local measurements in terms of spatio-temporal interest points. Such features capture local motion events in video and can be adapted to the size, the frequency and the velocity of moving patterns, hence, resulting in video representations that are stable with respect to corresponding transformations. In this work systematic study on human activity recognition using local invariant methods are described. Comparison of various studies understands that local feature is preserved for accurate recognition of human activity. The redundant information is removed during activity detection and classification of human activity frames also has been carried out by support vector machine (SVM) classifier.

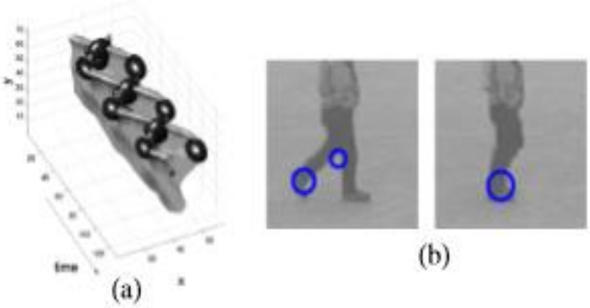


**Figure 1.** Local space-time features detected for a walking pattern :(a )3-D plot of a spatio-temporal leg motion (up side down) and corresponding features (in black). (b) Features overlaid on selected frame sequence.


**Figure 2.** Action and scenarios. Database (available on request): examples of sequences corresponding to different types of action.

## II. LITERATURE REVIEW

Ren and Xu (2002) [1] presented a new system for teachers' natural complex action recognition in the smart classroom in order to realize an intelligent cameraman and virtual mouse. First, the system proposes a hybrid human model and employs a second order B-spline function to detect the two shoulder joints in the silhouette image to obtain the basic motion features including the elbow angles, motion parameters of the face and two hands. Then, a primitive-based coupled hidden Markov model (PCHMM) is presented for natural context-dependent action recognition. Last, some comparison experiments show that PCHMM is better than the traditional HMM and coupled HMM.

Sathiya et al. [2] presented an important research area of computer vision. There is an urgent mechanism to automatically detect and recover semantic events in videos sequence contents. Low-level video sequence content is translated into high-level video sequence content. It is an application includes automated video surveillance schemes, intensive care system, airports, analysis of physical condition of people and variety of systems which include human computer interfaces. Automatic recognition of high level activities which refers combination of multiple simple human action based system.

Dharmsinh and Ratanpara [3], describes that human action recognition is a way of retrieving videos emerged from Content Based Video Retrieval(CBVR).Human action recognition has gained popularity because of its broad applicability in automatic retrieval of videos of particular action using visual features. The most common stages for action recognition includes: object and human segmentation, feature extraction, activity detection and classification. It introduces survey on different types of action like single person action, two people or person-object interaction and multiple people action recognition. Local space-time features detected for a walking pattern is shown in Figure 1.

Hajihashemi and Pakizeh [4] describes that Human activity recognition, based on fast growing the number of videos and multimedia data ,there is an massive demand for an accurate, fast algorithm that analyse, understand and clustered the content of these data. The speed and accuracy of these methods are so important. Another problem is the large data stored in each video that increase processing, waste time and memory. In this work a simple novel method for action recognition and video matching based on clustered codebook model of video frames are extracted. At second step, extracted features of video were clustered by K-means algorithm and five clusters for each video are made.

Yala et al. [5] describes that the human recognition system must analyse the changes in the sensing data and at any significant detection, it has to decide if there is a change in the activity performed by the person. Such a system can use the previous sensor readings for decision-making without the need to wait for future ones. This work proposes an approach of human activity recognition on online sensor data. Different action and scenarios are shown in Figure 2.

Anguita et al. [6] presented an Activity-Based Computing (ABC) aims to capture the state of the user and its environment by exploiting heterogeneous sensors in order to provide adaptation to exogenous computing resources. When heterogeneous sensors are attached to the subject’s body, they permit continuous monitoring of numerous physiological signals. This has appealing use in healthcare applications, e.g. the exploitation of Ambient Intelligence (AmI) in daily activity monitoring for elderly people. This work presents a system for human physical Activity Recognition (AR) using smart phone inertial sensors. This method adapts the standard Support Vector Machine (SVM).

Lara and Labrador(2013) [7], presented despite human activity recognition (HAR) being an active field for more than a decade, there are still key aspects that, if addressed, would constitute a significant turn in the way people interact with mobile devices. This work surveys the state of the art in HAR based on wearable sensors. A general architecture is first presented along with a description of the main components of any HAR system. Twenty eight systems are qualitatively evaluated in terms of recognition performance, energy consumption, obtrusiveness, and flexibility, among others. This work also, presents some open problems and ideas that, due to their high relevance, should be addressed in future research.

Arie et al. (2002) [8], develop a novel method for view-based recognition of human action/activity from videos. By observing just a few frames, they identified the activity that takes place in a video sequence. The basic idea of multidimensional indexing method is that activities can be positively identified from a sparsely sampled sequence of a few body poses acquired from videos. They developed a theoretical foundation that shows that robust recognition of a sequence of body pose vectors can be achieved by a method of indexing and sequencing and it requires only a few pose vectors.

Wang et al. (2007) [9], presented a human activity recognition based on a new feature descriptor. For a binary human silhouette, an extended radon transform, R transform, is employed to represent low-level features. The advantage of the R transform lies in its low computational complexity and geometric invariance. Then a set of HMMs based on the extracted features are trained to recognize activities. Compared with other commonly-used feature descriptors, R transform is robust to frame loss in video, disjoint silhouettes and holes in the shape, and thus achieves better performance in recognizing similar activities.

Dragan and Mocanu (2013) [10], presented a method for image based human activity recognition, in a smart environment. Author used background subtraction and skeletisation as image processing techniques, combined with artificial neural networks for human posture classification and hidden Markova models for activity interpretation. In this work defined approach is successfully recognized basic human action such as walking, rotating, sitting and bending up/down, lying and declining. The method can be applied in smart houses, for elderly people who live alone.

Kim et al. (2010) [11] presented an activity recognition can be exploited to great societal benefits, especially in real life, human centric application such as elder care and health care. Recognizing complex activity remains a challenging and activity area of research and the nature of human activities pose different challenges. The first focuses on accurate detection of human activities based on predefined activity model. An activity pattern discovery researcher builds a pervasive system first and then analyses the sensor data to discover activity patterns.

Gaglio et al. (2015) [12], presented a method for recognizing human activity using information sensed by a RGB-D camera, namely these of Microsoft Kinect. Experiments were performed on Kinect activity recognition dataset, a new dataset, and on CAD-60, a public dataset. The results shows the solution out performance four relevant works based on RGB-D image fusion, hierarchical maximum entropy Markova model, Markova random fields, Eigen joints.

Li et al. (2013) [13], presented a Recognizing and understanding the activities performed by people is a fundamental research topic in developing a wide range of applications that would be societal beneficial. In this, we present and discuss two research projects on human action recognition based on computer vision techniques. We also report an ongoing research project that focuses on learning human activities through low cost, unobtrusive radio frequency identification (RFID) technology.

Holte et al. (2012) [14], address the problem of human action recognition in reconstructed 3-D data acquired by multi-camera systems. We contribute to this field by introducing a novel 3-D action recognition approach based on detection of 4-D (3-D space + time) spatio-temporal interest points (STIPs) and local description of 3-D motion features. STIPs are detected in multi-view images and extended to 4-D using 3-D reconstructions of the actors and pixel-to-vertex correspondences of the multi-camera setup. Local 3-D motion descriptors, histogram of optical 3-D flow (HOF3D), are extracted from estimated 3-D optical flow in the neighbourhood of each 4-D STIP and made view-invariant. Experiments on the publicly available i3DPost and IXMAS datasets show promising state-of-the-art results and validate the performance and view-invariance of the approach.

Goudelis et al. (2017) [15], presented a Human action recognition is currently one of the hottest areas in pattern recognition and machine intelligence. Its applications vary from console and exertion gaming and human computer interaction to automated surveillance and assistive environments. We define the notion of a 3D form of the Trace transform on discrete volumes extracted from spatio-temporal image sequences. On a second level, we propose the combination of the novel transform, named 3D Cylindrical Trace Transform, with Selective Spatio-Temporal Interest Points, in a feature extraction scheme called Volumetric Triple Features, which manages to capture the valuable geometrical distribution of interest points in spatio-temporal sequences and to give reputation to their action-discriminate geometrical correlations. The technique provides noise robust, distortion invariant and temporally sensitive features for the classification of human actions

Gurukaynak and Yalcin (2015) [16], worked on a method that recognizes certain human activities based on a motion descriptor that uses 3D human skeleton data. A motion descriptor (SHOJD) is defined by the authors using the 3D distance between the most frequent key poses that occur throughout the action that is intended to be recognized. SHOJD features are then fed into an artificial neural network for classification. Their experimental results indicate that the SHOJD based human action recognition system is robust with high recognition rate.

Azim and Hemayed (2015) [17], developed a trajectory based local representation approaches to capture the temporal information. This work introduces an improvement of trajectory-based human action recognition approaches to capture discriminative temporal relationships. Authors extracted trajectories by tracking the detected spatio-temporal interest points named "cuboids features" with matching it's SIFT descriptors over the consecutive frames. They also, proposed a linking and exploring method to obtain efficient trajectories for motion representation in realistic conditions. Then the volumes around the trajectories' points were described to represent human actions based on the Bag-of-Words (BOW) model. Finally, a support vector machine has been used to classify human actions. The effectiveness of the proposed approach was evaluated on three popular datasets (KTH, Weizmann and UCF sports).

Zhang et al. (2014) [18], developed a novel action-scene model to learn contextual relationship between actions and scenes in realistic videos. With little prior knowledge on scene categories, a generative probabilistic framework is used for action inference from background directly based on visual words. Experimental results on a realistic video dataset validate the effectiveness of the action-scene model for action recognition from background settings. Extensive experiments were conducted on different feature extracted methods, and the results show the learned model has good robustness when the features are noisy.

Sun et al [19], developed most of the video action recognition use complex hand-designed local features, such as SIFT, HOG, SURE, but these approaches are implemented sophisticatedly and difficult to be extended to other sensor modalities, Recent studies discover that there are no universally best hand-engineered features for all datasets, and learning features directly from the data may be advantageous. SFA also observed that multi-layer features representation has succeeded remarkably in widespread machine learning application.

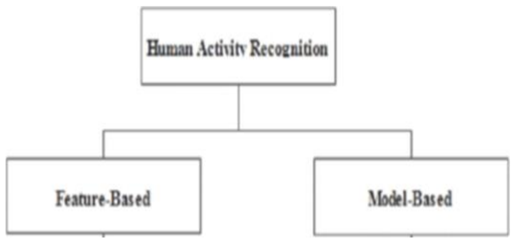


**Figure 3.** Overview of human activity recognition

Lin et al [20], developed a vision-based smart building control system relies upon human action recognition to determine the number of occupants inside the building and their respective motion type or path. Using this information control system that can automatically optimize environmental conditions within the building can designed. Figure 3 shows overview of human activity recognition.

## III. CONCLUSION

This survey has showcased the various Human Activity Recognition algorithms, various methodologies adapted for human object interaction in still images and videos, different human-human interaction techniques and types of classifications used for all these recognition. The goal is to provide an extensive survey and comparison of different techniques and approaches of Human Activity Recognition. Using this survey, some of the challenges and future directions are also highlighted. In summary, the literature in Human Activity Recognition depicts major progresses in various aspects. However these works still have not addressed the various challenges of activity recognition like incorporating the context of the scene, human-object interaction in videos, real-time activity prediction to its full extent.

## IV. REFERENCES

[1]. Haibing Ren, Guangyou Xu, "Human action recognition in smart classroom", IEEE International conference on Automatic Face and Gesture Recognition, 2002.

[2]. G. Akilandasowmya, P. Sathiya, P. Anandha Kumar "Human action recognition in research area of computer vision", IEEE International Conference on automatically detect and retrieve semantic events in video, 2015

[3]. Chandni J. Dharmsinh and Tushar V. Ratanpara "Human recognition in computer vision", IEEE International conference on automatic retrieval of videos of particular action using visual features Green engineering and Technology (IC-GET), 2016.

[4]. Vahid Hajihashemi and Esmat Pakizeh "Human activity recognition in videos based on a Two Levels K-means and Hierarchical Codebooks" International journal of Mechatronics, Electrical and Computer Technology, International conference on Oct 2016.

[5]. Nawel Yala, Belkacem Fergani and Anthony Fleury "Feature extraction for human activity recognition on streaming data" IEEE International Symposium on Innovation in intelligent Systems and Application (INISTA), 2015.

[6]. Davide Anguita, Alessandro Ghio, Luca Oneto, Xavier Parra, Jorge L. Reyes-Ortiz " Human Activity Recognition on Smart phones Using a Multiclass Hardware-Friendly Support Vector Machine" volume 7657.

[7]. Oscar D. Lara and Miguel A. Labrador "A Survey on Human Activity Recognition using Wearable Sensors" IEEE Communications Surveys and Tutorials, Volume: 15, Issue: 3, Third Quarter 2013.

[8]. J. Ben-Arie, Zhiqian Wang, P. Pandit, S. Rajaram "Human activity recognition using multidimensional indexing" IEEE Transactions on Pattern Analysis and Machine Intelligence, Volume: 24, Issue: 8, 2002.

[9]. Ying Wang, Kaiqi Huang, Tieniu Tan "Human Activity Recognition Based on R Transform" IEEE Conference on Computer Vision and Pattern Recognition, 2007.

[10]. Monica-Andreea Dragan and Irina Mocanu "Human activity recognition in smart environment" IEEE International conference on Control System and Computer Science, 2013.

[11]. Eunju Kim, Sumi Helal, Diane Cook "Human Activity Recognition and Pattern Discovery" IEEE transaction on Pervasive Computing, Volume: 9, Issue: 1, 2010.

[12]. Salvatore Gaglio, Giuseppe Lo Re, Marco Morana "Human Activity Recognition Process using 3-D posture data" IEEE Transactions on Human-Machine Systems, Volume: 45, Issue: 5, 2015.

[13]. Xue Li, Quan Z. Sheng, Chaoyi Pang, Xin Zhao, Sen Wang "Effective approaches in human action recognition" International Conference on Advanced Computer Science and Information System (ICACSIS), 2013.

[14]. Holte, M., B. Chakraborty, J. Gonzalez, and T. Moeslund, "A Local 3-D Motion Descriptor for Multi-View Human Action Recognition from 4-D Spatio-Temporal Interest Points" IEEE Journal in Signal Processing, Volume: 6, Issue: 5, Sept. 2012.

[15]. Georgios Goudelis, Georgios Tsatiris, Kostas, Karpouzis, Stefanos Kollias "3D cylindrical trace transform based feature extraction for effective human action classification" IEEE Conference on Computational Intelligence and Games (CIG), 2017.

[16]. Deniz Gurkaynak , Hulya Yalcin, "Recognition and classification of human activity", IEEE conference on Signal Processing and Communications Applications, 2015, ISSN: 2165-06008.

[17]. Haiam A. Abdul-Azim, Elsayed E. Hemayed, "Human action recognition using trajectory-based representation", Egyptian Informatics Journal, Volume 16, Issue 2, 2015, pp. 187-198.

[18]. Yifei Zhang, Wen Qu, Daling Wang Action-scene Model for Human Action Recognition from Videos AASRI Precede, Volume 6, 2014, pp. 111-117.

[19]. Lin Sun, Kui Jia, Tsung-Han Chan, Yuqiang Fang, Gang Wang, Shuicheng Yan "Human recognition in complex hand-designed local features", IEEE conference on Computer Vision and Pattern Recognition 2 014.

[20]. S. C. F. Lin, C. Y. Wong, T. R. Ren and N. M. Kwok "Human action recognition in smart building control system" IEEE International Conference on automatically optimize environmental conditions within building. 2012