Insights on Research-based Approaches in Human Activity Recognition System

Abdul Lateef Haroon P. S.
Department of Electronics & Communication Engineering,
Ballari Institute of Technology and Management,
Karnataka, India

U. Eranna
Department of Electronics & Communication Engineering,
Ballari Institute of Technology and Management,
Karnataka, India

ABSTRACT

There has been increased proliferation of Human Activity Recognition system to be embedded in the different form of sensing technologies. With the faster advancement of novel features in the sensory application, the human activity can be used as a tool to either command the system from remote or could be used to perform sophisticated analysis of human behavior. Since last decade, there has been the volume of literature focusing on leveraging the identification process using different forms of research-based methodologies. However, it is quite evident that there is no benchmarked model and nor a signatory research work in this field that has been observed till date to be standardized among the research community. Hence, this paper investigates the fundamentals, different existing approaches, and loopholes associated with such approaches so that potential and impending problems associated with it can be distinctively explored. The paper contributes to the identification of some of the open research issues which need significant attention.

Keywords

Human Activity Recognition, Motion Sensing, Action, identification Accuracy

1. INTRODUCTION

Generally, human beings can recognize and understand their actions clearly by visual information. In our lives could be revolutionized if computer/machines can automatically understand & recognize the human activities. Thus, Recognition of Human Activities (RHA) has increasing interest to the researchers in the communication computing world [1,2,3 and 4]. The primary objective is to analyze and predict the human actions and to perceive different types of activities. Based on any one of RGB-data, skeletal joints, depth-maps, and incorporation of these modalities, several activity recognition methods have been introduced and utilized in a specific application, e.g., the interaction of human computer, video surveillance, military, medical applications so on [5]. The human actions recognition process can be performed at different levels of abstraction. Similarities occur between different taxonomies like gesture, motion, activity. Based on the study [3], [6]. "Gestures" are the elementary movements of the human body. Generally, they are initially definable components which can be defined at the level of body-parts (e.g., leg kick and arm stretch). While "Movements" are the single person activities, containing several gestures which have a temporal ordering. They are distinguished as entire body movements (for example walking, running, sitting, etc.). "Activities" refer to the multiple numbers of subsequent actions/movements often performed by multiple persons, with or without objects for example "playing cricket" or "operating a computer." However, there are several RHA approaches have been introduced, but there are still many challenges towards the implementing of new methods to enhance the accuracy under realistic scenarios. Such challenges are (i) election of suitable attribute to be measured, (ii) to build a portable, non-obtrusive and cost effective recognition system, (iii) design of feature extraction methods (iv) data incorporation under realistic scenario, and (v) development of smart device to achieve processing and energy requirements. The RHA has been approached in two ways, viz using External devices and Wearable sensor devices. In a conventional approach, the systems are fixed on the pre-determined point of interest. Thus all the inference activities depend upon user's interaction with sensors. Later, devices are attached to the person. Intelligent-Homes are the most common example for external sensor devices. These sensors are capable to fairly recognize the complex activities like eating, washing, taking a shower and so on. Since they rely on information from sensors which are fixed in target objects and humans can interact with objects (stove, washing machine, and faucet, etc.). Moreover, the installation and management of this kind of sensor devices usually available at high costs. To overcome these challenges, researchers found the use of wearable sensor devices in the recognition system. Almost every measured attribute belong to person activity (using GPS and accelerometers), environmental attributes (like humidity and temperature), and physiological signals (measuring heart rate or electrocardiogram.). This information is basically indexed above the time period and provides us to recognize the human activities. In past decades, the study on RHA has mainly focused on learning and recognizing person activities from the video sequences captured from traditional cameras [1], [4], [7]. Captured video from traditional cameras encodes high texture with color information, which is helpful for image processing. One of the most significant challenges for the researchers is that capturing the 3-D action from traditional cameras. As human activities are performed in 3-D space, accessing 3-D information plays a major role in activity recognition process. However, Depth-Map approach is one of the best techniques utilizing for successful human activity recognition system [8]. Compared with the traditional approaches depth-map approach shown several benefits in the context of activity recognition process. For example, this method can provide 3-D structural information. The primary contribution of this survey study is to offer the complete view of the methodology of human activity recognition using sensors. Section-2 discusses generate c structure of RAH system followed by various design challenges over RAH system in section-3. Section-4 highlights the different existing approaches. Finally, the last section discusses open research challenges in the area providing directions for future study.

2. 2. ARCHITECTURE OF RECOGNITION OF HUMAN ACTIVITIES

The architecture of any Recognition of Human Activity (RHA) system based on the activities to be recognized. Actually, changing of human activities immediately turns into to RHA problem. Form the survey study, human activities can be distinguished into different categories and each activity belongs to individual category, e.g., 1) Ambulation (Walking, Running, and Lying, Climbing stairs), 2) Transportation (Riding a bus, cycling), 3) Daily Activities (Eating, drinking, working, etc.). However, the RHA system contains two significant phases, i.e., the Training phase and testing phase. The following figure-1 represents the generic design of the RHA system.

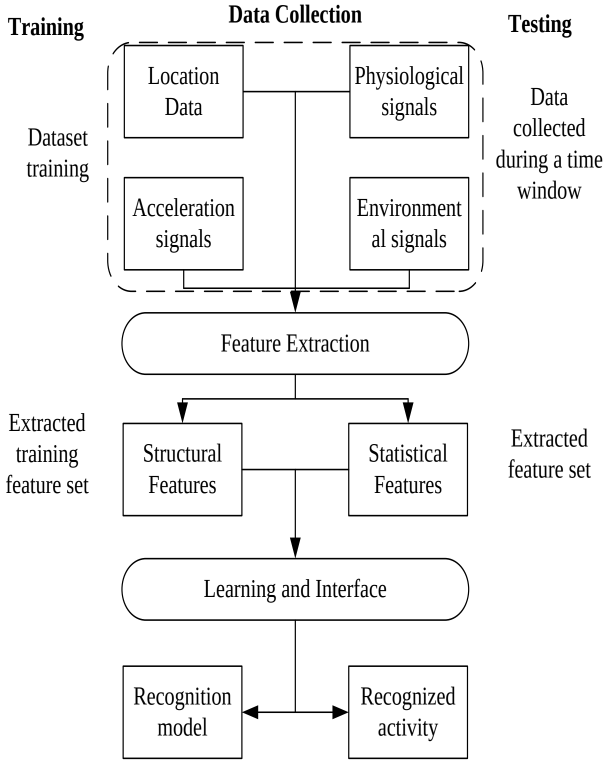

Fig 1: Generic Structure of RHA System

Initially, the training phase needed a time series of a dataset and measured the attributes by performing each human activity. The time series datasets are divided regarding time windows which apply for feature extraction. After that, learning methods are utilized to design a human activity recognition model from the extraction of dataset features. Similarly, in a testing phase, data will be collected during a time window and utilized for feature extraction. These feature sets are computed by existing trained learning method, creating a predictive activity model.

3. 3. DESIGN CHALLENGES IN RHA SYSTEM

This section briefly discusses significant design challenges about RHA. Namely, i) non-obtrusive, ii) selection of attributes, iii) energy consumption, iv) performance recognition, v) processing Speed.

i) Non-obtrusive

By the successful implementation, the RHA system does not need the person to wear a number of sensors nor interact too often with the application. Some existing human activity recognition models require the user to wear more than four accelerometers or to carry heavy recording devices [9] [10]. These recognition devices may be uncomfortable, expensive or invasive; therefore these are not suitable for human action recognition. Other models introduced [11], [12] which recognize the human activities by cellular devices.

ii) Selection of Suitable Attributes

There are some important attributes evaluated by exploiting wearable sensor devices such as; Location, Acceleration, Environmental attributes (e.g., temperature, humidity, etc.) and physiological signals. Environmental attributes are intended to give context information about surrounding environment. Whereas, triaxial-accelerometers can recognize the ambulation activities (e.g., walking, running, etc.) but these devices are less significant to recognized with high accuracy. For instance, accelerometer sensor on the wrist may not recognize the human activities efficiently, because accidental movements can produce incorrect predictions. However, to measure the physiological signals, external sensor devices can be required and these sensor made up with the wireless system which entails high energy expenditures.

iii) Energy Consumption

Most of the context-aware applications rely on smart devices like mobile phones, which are basically energy constrained. In most of the cases, saving the battery life is a significant feature, particularly for clinical and defense applications which are compelled to forward significant information. Unexpectedly, most of the activity recognition methods do not accurately analyze the energy expenditure, which is very essential due to computation, processing, and for task visualization.

iv) Processing Speed

Measuring of activity recognition in mobile/sensor devices becoming more challenging because they are still constrained in terms of storage, energy consumption, and processing. For instance, some learning algorithms are very costly during experimental phase, which creates them not efficient for sensor human activity recognition.

v) Performance Recognition

The overall performance of the activity recognition system considers various aspects, e.g., set of activities, quality of the training data set, feature extraction technique and learning algorithm. Basically, every individual activity brings a different pattern of recognition problem. Hence, it is essential to understand the performance of activity recognition using some standard methods for example: F-Measure, ROC curves, and Kappa Statistic [13].

4. 4. EXISTING APPROACHES OF RHA

In section-2, have already discussed generithe c structure of the recognition of human activities, basically collected data have to pass through the process of feature-extraction. Later, the recognition system is created from the set of extracted features via machine learning approach. Once the training process completed, unseen instances (time-windows) can be measured in the recognition model, resulting from a prediction upon performed actions. Thus, feature extraction and machine learning are the two main approaches which play a significant role in the RHA process.

4.1 Feature Extraction Approach

Human performed the different activities during a long period compared with sensors sampling rate (~ 250 Hz). Besides, one sample in a particular instance of time will not give enough information about the performed activity. Therefore, for activity recognition will exploit time-windows basis method rather that sample basis. But, the challenge is; how do we distinguish two given time windows? It may not be possible for the signals to be precisely measured. So this is the primary reason for adopting feature extraction method to the individual time window, which relevantly filters the data and generates the quantitative results. Basically, there are two significant approaches, which extract the features from the time series of data such as i) Structural approach and ii) Statistical approach [14]. The Statistical approaches, e.g., Wavelet transform & Fourier transform, exploit quantitative data characteristics to extract the features, while structural approach considers into account the interrelationship between the series of data.

4.2 Machine-Learning Approach

Researchers introduced some important machine learning tools, which helps to recognize the different activities as well as analyze and predict the data also. In this approach, the data patterns are to be developed from a set of data or instances, and such input datasets are named as training data sets. Here, each data set is feature vector and extracted by signals within a time window. However, there are two exiting machine learning methods, viz supervised learning method and un-supervised learning method which are deals with labeled and un-labeled data respectively. Most of the RHA systems adopt the supervised learning approach which provides the efficient results.

However, WEKA (i.e., Waikato Environment for Knowledge Analysis) tool is a most efficient tool utilized in a machine learning research field [15]. It involves computation of a number of learning algorithms, and it offers to simplify them easily for specific data set using cross evaluation and random split. The main advantage is that it provides a JAVA platform which provides the incorporation of advanced learning algorithms and computation methods on the top of the pre-existing framework. One of the drawbacks of the WEKA tool is that it is not fully functional in present mobile platforms. Thus the following figure highlights the most significant learning algorithms which are majorly utilizing in recognition of human activity process.

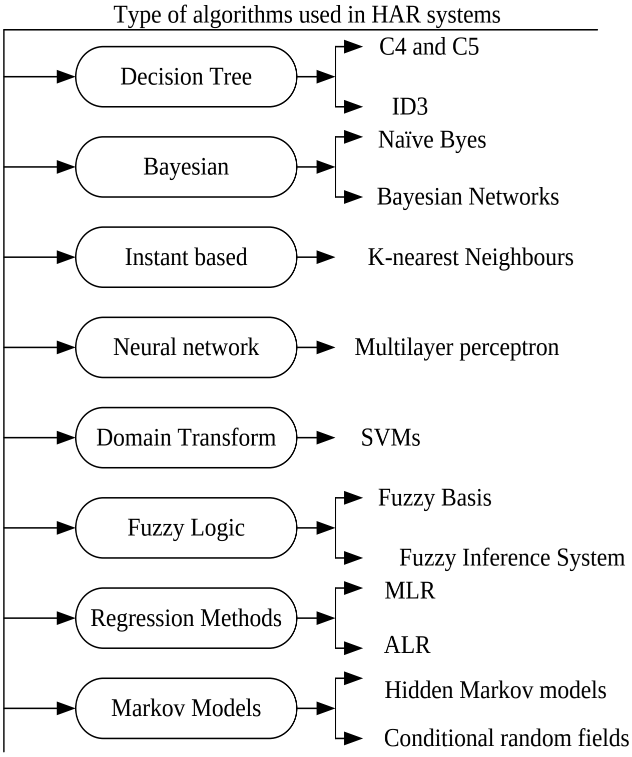

Fig 2: Types of Learning Algorithms used in HAR System

5. 5. MODALITIES OF DATA

Depth-Maps (DM) and Skeleton Joints (SJ) are the two most common data-modalities which are generally used in action recognition system. The DM and SJ not only provide a successful human action capturing method but also make it easy to design human activity recognition model effectively.

5.1 Depth Sensors

Depth sensors can sense the 3-D visual-world and collect the lowest level visual information. In DM, the depth silhouettes object can be extracted very easily and precisely. Based on the modalities of depth-sensors, different remarkable feature-extraction and representation methods have been introduced like DM-based space-time features and DM-based sequential features.

5.2 Skeleton Joints (SJ)

It encodes the 3-D human skeletal joint positions for an individual frame in real time. Every movement of the skeleton can be differentiating as several actions; using skeleton data for movement recognition is a potential direction. Several recognition algorithms have been introduced and utilized to design a skeleton from depth maps [16], [17]. The idea behind these algorithms is to divide the depth information of human body into several parts with dense-probabilistic labeling. The body part segmentation can be taken for the task of classification for the single pixel in depth-maps. Based on the modalities of depth-sensors, different remarkable feature-extraction and representation methods have been introduced like Skeletal-based space-time features and Skeletal-based sequential features. Generally, skeletal based features are simpler to extract and need very less memory and computational cost compared with DM based feature.

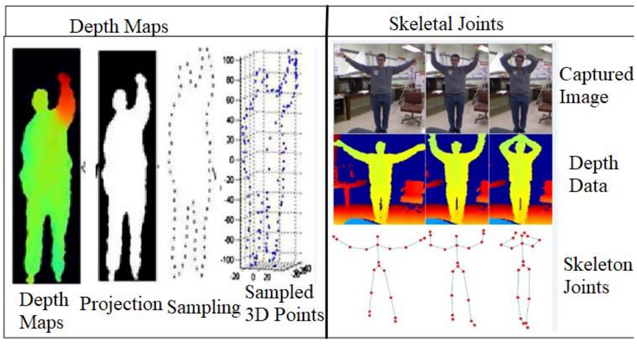

Fig 3: Comparative analysis of Depth Maps and Skeleton joints

The above figure-3 represents the comparative analysis between DM and SJ data modalities.

6. 6. EXISTING APPROACHES

This section discusses the existing approaches of implementing various problems associated with human activity recognition. The discussion carried out in this section was extracted from the research papers in IEEE digital library from 2010 till 25th Feb 2018. Though the keyword search "Human activity recognition" index approximately 92 journals but still some papers are not of much relevance.

The very objective of recognition of human activity is to correlate it as a context for ubiquitous applications. The use of wearable sensor was adopted to recognize the human activities. In the work of Nishimura et al. [18], the authors have proposed a method for the recognition by the pattern from multi models of images, sound and acceleration signals. Though they argue a low computational overhead, these frameworks lack the accurate result against real-time challenges. The accuracy can be improvised if the neighborhood decision is clear and accurate. One of the works by Ribeiro et al. [19], the fuzzy is used for sampling of the neighboring to increase the robustness, and when it is applied on the human activity recognition and object detection problem, it exhibits better result. The HAR based on the sensor reading poses power constraints due to the limited battery. Therefore a combined optimization problem for both user state recognition and power optimization is studied by Yurur et al. [20], by using Hidden Markov Model (HMM) and learning from data concepts. The problem of HAR becomes more challenging if it moves from labeled to the non-labeled domain that requires a continuous learning mechanism and even the deep-learning based model is unsuitable for continuous learning due to large network constraints of it. In the work of Hasan et al. [21], an activity learning method is proposed to select most suitable feature automatically, and the model is validated for four different human activity datasets to check its effectiveness. In the domain of HAR, locating 3D joints plays an important role.

The effective use of depth sensors is explored for the purpose of full body tracking in the real-time scenario that requires accurate identification of skeletal. In the work of Li et al. [22] have addressed the problem related to the recognition of human skeletal from the data captured in different views from the depth sensors using graph model. Yet another approach for the activity recognition is adopted by Huang et al. [23] based on the wireless signal especially for the indoor environment. It involves the challenges of handling non-line of sight, velocity of human etc. For interpretation of data by human a proper classification and clustering is require to categorize. The use of hidden Markov model is done by Troelsgaard et al. [24] which used a concept of score function, their method provisions lower computational complexity for classification time and the method is validated for human activity recognition. The closely associated skeleton based human activity recognition is studied. Initially, due to the high cost of the depth camera the human activity/ behaviour methods considers only the x-y and time as an input vector in respective methodologies. These methods losses the discriminative features as it do not consider the depth information. The recent advancement into depth sensors has brought focus of researchers on consideration of the depth axis also along with spatial-temporal axis for effective HAR. Many methods have considered RGB-D dataset for the study, these studies include [25, 26, 27, 28 and 29], where it is realized that the depth information is very useful cue in the appearance of the human activities. In the work of Yang et al. [30], the investigation was carried out for HAR in the absence of skeleton for overcoming the constraints of line of sight alignment with depth sensor camera and assumption of no occlusion by means of interest point extraction and depth map descriptor. They validated their method for the senior activity recognition. The World Health Care Organization (WHO) has produced a report for the statistics and requirements for the fall presentations of old aged people [31]. The work of Withanage et al. [32], it is discussed that due to the accidental falls of elderly aged people many mortalities takes place across the world. This requirement leads to have a proper fall detection method that may include a in-situ robot guided support system on the basis of many behaviour such as falling, rolling, moving on hands and knees, crawling, etc. kinds of human activities/ behaviour recognition. They use low level image features along with depth to achieve higher accuracy as compared to the nearest competitor. Another work where a depth camera is used for the HAR is developed by Yang et al. [33]. In their method They have extended the normal surface to the polynomial by combining the local neighboring hyper-surface normal from a depth sequence to jointly characterize local motion and shape information and further a concept of super normal vector (SNV) is used to aggregate the low-level polynomial into a discriminative representation, which can be viewed as a simplified version of the Fisher kernel representation. They validated their model for the superior performance as compared to the state of the artwork on four different public domain dataset that includes 1) MSRAction3D, 2) MSRDailyActivity3D, 3) MSRGesture3D and MSRActionPairs3D. One of the very recent works towards elderly monitoring system is considered for the in-depth study as well as a state of artwork for the further optimization as a contribution to this research problem domain is done by Hbali et al. [34], where smart environment in the home is advocates for the aging people health care. In their work they have used 3-D depth sensors and proposed a method which is skeleton based and using Minkowski and Cosine distance for finding the accurate 3D joints.

Another scope of these research leads to the domain of Ambient Assisted Living (AAL), where the capability to recognize human behaviour and the interaction between two (or more) persons plays an important role. In the work of Manzi et al. [35], has considered two-person activity recognition using the skeleton data extraction from the depth camera. They use unsupervised clustering approach by encoding the human actions as a basic posture. The model is validated on two datasets namely 1) Institute of Systems and Robotics (ISR) - University of Lincoln (UoL) and 2) Stony Brook University (SBU) datasets, and archives overall accuracies of 0.87 and 0.88, respectively. Ultimately, the very core research issue is to develop the mechanism for the highly accurate identification of the human action or behaviour in the very low overhead of computational complexities so that it can be used into applications like finding the abnormalities actions of human behaviour [36]. Yet another scope of an application called user identification based on the human behaviour recognition [37].

The important contributions/work done in the directions of human activity/ behavior recognition system on the basis of only RGB content is studied as follows: The work carried out by Luo et al. [38] have jointly used temporal-dynamic as well as feature-based approach to address the problem of human activity behaviour. The authors have also used learning-based approach for further improving the recognition outcome. Another work of Shan et al. [39] has constructed the slice-based approach to obtain optimal slice coordinates to improve the human behavior sequence into spatiotemporal space. In the study of Guo and Chen [40], the authors have addressed the problem of feature descriptor and single-task learning technique in human activity recognition and offered a novel multi-classification learning approach based on visual characteristics of the human structure. With another approach for cross-view activity recognition, Zheng et al. [41] have presented multi-dictionary learning method in which first it performs the specific task to analyze the angle of each view and secondly, it formulates a featured based model which uses the common dictionary that having records of different views. The outcome of this study shows that it had good ability that understands an action from unseen views. In the same way the work of Yuan et al. [42], the authors have used the multiple techniques which including R-features transform with the context-aware kernel learning algorithm that captures the distributed actions and to classify the similarities between the representations of activity in the video sequence. The experimental performance delivers that the proposed approach is effective for human action recognition from the captured videos.

The study of Soomro et al. [43] and Kuehne et al. [44] has conducted an activity recognition performance with very large data set from much different localization. Similarly, Liu et al. [45] have introduced the multi-dimensional interactive human activity recognition Data-set model with including four learning approaches that are cross-view learning, cross-domain learning, single-view learning, and multitask learning and in which all the actions are captured with unconstrained body movement. Liu et al. [46] have introduced the genetic programming algorithm with adaptive spatiotemporal descriptors which delivers the optimal way to fuse the color and movement of action into a single representation form to perform the action recognition task without any profound knowledge of sampled activity data-sets. In the study of Liu et al. [47], the authors have offered clustering learning technique for combined human action Recognition. The presented work is performed into three tasks are action grouping, action recognition and comparison between proposed technique and multi-task cluster learning approach. Wu et al. [48] have studied various Recognize Actions learning approach and suggested new learning approaches based on Fisher vector and vectors of locally aggregated descriptors for human action recognition. In work of Peng et al. [49] has evaluated performances of various realistic datasets with Bag of Visual Words model which formulates global representation by a set of local features which includes four steps such as codebook generation, feature encoding, feature extraction and pooling and normalization.

With the invent of depth sensor, the depth information's are obtained, and a distinguished dataset namely RGBD is obtained that can improvise the accuracy of the human activity/ behaviour recognition capacity as well the RGBD can facilitate the estimation of the human Skelton. The important contributions/work done in the directions of human activity/behaviour recognition system on the basis of only RGBD is presented as follows: Some other studies have also considered body structure pattern such as pose, body shape, and clothing for predicting the action from single depth image, these studies carried out by [50][51], where the transitional representation model is constructed so that each pixel image is classified into simpler form that locates each joint of body. The significant work towards action recognition is presented by Xia et al. [52] in which a visual of steady representation of human skeleton is formulated by using a feature of 3D Joint Locations in histogram based on improved spherical coordinates. The outcome of this study display that presented approach achieves real-time performances. Yang and Tian [53] have proposed a modified spatial, temporal pyramid which captures globally spatial orders and temporal order and another method for collective low-level polynomials into a supernormal vector. The effective outcomes presented approach obtains superior classified results compared to some public datasets.

In the study of Rahmani et al. [54], the authors have worked of 3D action recognition for which the author has used the descriptor with a concept of Histogram components and detection algorithm to find Space-time key factor in 3D point-cloud sequences. Yang et al. [55] have formulated Eigen-Joints concept by skeleton action data which applies to know the position differences between the joints that help to present human action and also, Naive-Bayes nearest neighbor with the principal component analysis is used for action classification. An action-let based approach is used in work of Wang et al. [56] for capturing intraclass variance with minimum noise and error-free in joint positions that perform human action recognition by a depth camera. In the study of Ofli et al. [57], the authors gave a new representation approach for the Human skeleton action recognition. The author has presented informative joint sequence in which the skeleton joints were automatically selected by extremely interpretable measures as like mean of joint angles, an angular velocity of joints. The objective of the proposed method is to achieve better human action recognition task. The study of Li et al. [58] has selected an action graph model and bag of 3D points to action model which performs the task of human activity recognition from the depth images. Further, one of the significant problems which are being identified by WHO called fall detection is studies and has attracted the focus of the research community in the recent past. The method adopted is classified into three categories 1) 3D skeleton-based approaches, 2) Depth-based approaches and 3) Hybrid approach. In the work of Zanfir et al. [59], the authors have tried to mitigate the problem of inappropriate segment action sequences by introducing a robust and efficient non-parametric descriptor of moving pose for low-latency activity recognition task. Tamou et al. [60] have offered a new descriptor for action recognition based on the 3D-differences of skeleton joints extracted from RGB-D cameras, and Random Forest classifier is utilized for action classification task. In the work of Keceli and Can [61], the authors have used depth data from Kinect sensor to perform basic task human action recognition, and the action classification is done through random forest algorithm and support vector machines. The presented approach is tested on the various 3D action data sets such as HUN-3D, MSRC-12, and MSR which results in achieving efficient human action recognition task. A new effective representation of Skeleton with low latency human activity recognition is presented by the Cai et al. [62] in which multi-channel with multiple instance learning and Markov random field is used. The author has suggested that the presented method is very effective controlling un-segmented sequences.

The work carried out by the Hussein et al. [63] has reported the problem of sequence representation of Skelton joint motion. The author has presented a new descriptor depend on the hierarchy of covariance matrix of skeleton joints coordinates. In the study of Xia and Aggarwal [64], the authors have proposed an algorithm for the space-time interest point that tackles the problem of noise in the depth data and uses a descriptor for 3D cuboids in-depth videos and lastly, the author combined both techniques that perform the action recognition task from depth videos in order to achieve efficient human activity recognition also more applicable than other existing models. The work of Vieira et al. [65], the authors gave a new representation approach that is Spacetime Occupancy Patterns for 3D action recognition in which the time and space axes are divided into the cells to represent the depth sequences. Oreifej et al. [66] have introduced depth-based descriptor by using a histogram to record the normal surface in the 4d volume depth with spatial coordinates. The work of Yang et al. [67] has presented action recognition method by formulating Depth Motion Maps. And a Oriented Gradients Histograms are used to captures the all activity from a side and front view. Zhao et al. [68] have presented a optimal method for human activity recognition by using the combination of RGB and depth map features. Amor et al. [69] have focused on classifying action by depth sensors using median filtering and downsampling. Yu et al. [70] presented a mechanism for real-time action recognition. Chaaraoui et al. [71] have constructed a joint approach based on skeletal feature and silhouette feature to obtain efficient visual characteristics and improved human action recognition.

7. OPEN RESEARCH ISSUES

This section discusses the open issues associated with approaches introduced by existing system in the prior section.

7.1 Less Focus on Joint Identification

A closer look at the existing system highlights that majority of the existing system to directly take the feed of the input image from the dataset followed by subjecting the processed image into its respective algorithm. However, it has not found that much work is carried out towards identifying the significant joints in the skeleton-based modalities, which is one of the essential driving forces to ensure accuracy in classification process in later part of the operation in the human activity recognition system.

7.2 The inclusion of Iterative-based process

Existing solutions towards human activity recognition system are more inclined towards using iterative-based process, which performs lots of unnecessary and repetitive number of operations. It will directly mean that existing mechanism has no emphasis on computational performance and is only concern about the identification accuracy. There is the inclusion of various training-based approaches as well as classifier used in the majority of existing approaches which uses an increasing number of epoch to obtain the better outcome.

7.3 Less novelty in approaches

The mechanisms adopted by the existing system of human activity recognition system are nearly similar in the majority of the approaches. There is an inclusion of three stages of operation, i.e., aggregation of the signal, extraction of the feature, and learning process followed by inference of it. The process of data collection is accompanied by both pieces of training as well as a testing process for ensuring better feature extraction process. However, there was no novelty explored in highlighting significant classification techniques apart from this. Every existing standard system is associated with advantages and limitations, and an effective way to explore novelty is by applying a hybridized form of classification techniques, which doesn't exist at present.

7.4 No Practical Evidence of real-time Utility

Majority of the existing mechanism of activity recognition system depends on three-dimensional datasets of actions which has a lesser scope of real-time utilization. It is because in real-time, it is not necessary that recognition should be carried out from single window operation that captures the feed of the object. With the principle of homography, it is highly possible that one object or multiple objects be sensed from multiple visual sensors. In such case, none of the existing approaches will be the application as they have reduced the scale of the temporal factor associated with it. There is a need of developing completely a new optimized framework that can offer identification concerning multiple visual feeds.

All the above points relate to the research gap that is immensely required to be addressed to further leverage the real-time utilization of human activity recognition system to scale up with future applications.

8. CONCLUSION

Sensing technology is going to be revolutionized in very near future, and hence it will increase potentially upgrade the performance of human activity recognition system. At present, there are various research-based approaches that have been addressing different forms of problems associated with action, activity, or gestures differently. Usage of training-based approach is found to be a mandatory requirement of such application design. However, one of the potential pitfalls in this process is that training operation is majorly carried out before feature extraction process. This process results in the further deterministic selection of all the feature-based information even if they are actually not required. Moreover, the performance could significantly differ from one to another database. After reviewing the existing literature, it can be found that existing approaches offers a comprehensive guideline to carry out future research work, but at the same time, there are some significant pitfalls too as identified in the prior section. Hence, the future work will be towards the direction of overcoming such problems. The first order of future work should be focused on identification of the potential information to address the problem of unnecessary information gathering process. The second order to future work will be focused on improving the accuracy of the recognition system while the third order to future work will be focused on investigation towards hybridizing the existing system for the effective classification process.

9. REFERENCES

[1] T. B. Moeslund, A. Hilton, and V. Kr ̈uger, "A survey of advances in vision-based human motion capture and analysis," Computer vision and image understanding, vol. 104, no. 2, pp. 90–126, 2006.

[2] S. Mitra and T. Acharya, "Gesture recognition: A survey," Systems, Man, and Cybernetics, Part C: Applications and Reviews, IEEE Transactions on, vol. 37, no. 3, pp. 311–324, 2007.

[3] P. Turaga, R. Chellappa, V. S. Subrahmanian, and O. Udrea, "Machine recognition of human activities: A survey," Circuits and Systems for Video Technology, IEEE Transactions on, vol. 18, no. 11, pp. 1473– 1488, 2008.

[4] R. Poppe, “A survey on vision-based human action recognition,” Image and vision computing, vol. 28, no. 6, pp. 976–990, 2010.

[5] A. Jaimes and N. Sebe, “Multimodal human–computer interaction: A survey,” Computer vision and image understanding, vol. 108, no. 1, pp. 116–134, 2007.

[6] J. K. Aggarwal and M. S. Ryoo, “Human activity analysis: A review,” ACM Computing Surveys (CSUR), vol. 43, no. 3, p. 16, 2011.

[7] D. Weinland, R. Ronfard, and E. Boyer, “A survey of vision-based methods for action representation, segmentation and recognition,” Computer Vision and Image Understanding, vol. 115, no. 2, pp. 224–241, 2011.

[8] A. Janoch, S. Karayev, Y. Jia, J. T. Barron, M. Fritz, K. Saenko, and T. Darrell, “A category-level 3d object dataset: Putting the kinect to work,” in Consumer Depth Cameras for Computer Vision. Springer, 2013, pp. 141–165.

[9] J. Parkka, M. Ermes, P. Korpipaa, J. Mantyjarvi, J. Peltola, and I. Korhonen, “Activity classification using realistic data from wearable sensors,” IEEE Trans. Inf. Technol. Biomed., vol. 10, no. 1, pp. 119– 128, 2006.

[10] U. Maurer, A. Smailagic, D. P. Siewiorek, and M. Deisher, “Activity recognition and monitoring using multiple sensors on different body positions,” in Proc. International Workshop on Wearable and Implantable Body Sensor Networks, (Washington, DC, USA), IEEE Computer Society, 2006.

[11] M. Berchtold, M. Budde, D. Gordon, H. Schmidtke, and M. Beigl, “Actiserv: Activity recognition service for mobile phones,” in International Symposium on Wearable Computers, pp. 1–8, 2010.

[12] S. Reddy, M. Mun, J. Burke, D. Estrin, M. Hansen, and M. Srivastava, “Using mobile phones to determine transportation modes,” ACM Trans. Sensor Networks, vol. 6, no. 2, pp. 1–27, 2010.

[13] D. Lara, Oscar & Labrador, Miguel. (2013). A Survey on Human Activity Recognition Using Wearable Sensors. Communications Surveys & Tutorials, IEEE. 15. 1192-1209. 10.1109/SURV.2012.110112.00192.

[14] R. T. Olszewski, C. Faloutsos, and D. B. Dot, Generalized Feature Extraction for Structural Pattern Recognition in Time-Series Data. 2001.

[15] “The Waikato Environment for Knowledge Analysis,” http://www.cs. waikato.ac.nz/ml/weka/.

[16] R. Girshick, J. Shotton, P. Kohli, A. Criminisi, and A. Fitzgibbon, “Efficient regression of general-activity human poses from depth images,” in Computer Vision (ICCV), 2011 IEEE International Conference on. IEEE, 2011, pp. 415–422.

[17] J. Shotton, T. Sharp, A. Kipman, A. Fitzgibbon, M. Finocchio, A. Blake, M. Cook, and R. Moore, “Real-time human pose recognition in parts from single depth images,” Communications of the ACM, vol. 56, no. 1, pp. 116–124, 2013.

[18] J. Nishimura and T. Kuroda, "Multiaxial Haar-Like Feature and Compact Cascaded Classifier for Versatile Recognition," in *IEEE Sensors Journal*, vol. 10, no. 11, pp. 1786-1795, Nov. 2010.

[19] P. C. Ribeiro, P. Moreno and J. Santos-Victor, "Introducing fuzzy decision stumps in boosting through the notion of neighbourhood," in *IET Computer Vision*, vol. 6, no. 3, pp. 214-223, May 2012.

[20] O. Yurur, M. Labrador and W. Moreno, "Adaptive and Energy Efficient Context Representation Framework in Mobile Sensing," in *IEEE Transactions on Mobile Computing*, vol. 13, no. 8, pp. 1681-1693, Aug. 2014.

[21] M. Hasan and A. K. Roy-Chowdhury, "A Continuous Learning Framework for Activity Recognition Using Deep Hybrid Feature Models," in *IEEE Transactions on Multimedia*, vol. 17, no. 11, pp. 1909-1922, Nov. 2015.

[22] M. Li and H. Leung, "Multiview Skeletal Interaction Recognition Using Active Joint Interaction Graph," in *IEEE Transactions on Multimedia*, vol. 18, no. 11, pp. 2293-2302, Nov. 2016.

[23] X. Huang and M. Dai, "Indoor Device-Free Activity Recognition Based on Radio Signal," in *IEEE Transactions on Vehicular Technology*, vol. 66, no. 6, pp. 5316-5329, June 2017.

[24] R. Troelsgaard and L. K. Hansen, "Sequence Classification Using Third-Order Moments," in *Neural Computation*, vol. 30, no. 1, pp. 216-236, Jan. 2018.

[25] Shotton J, Fitzgibbon A, Cook M, et al. Real-Time Human Pose Recognition in Parts from Single Depth Images[C]// Proceedings of IEEE Conference on Computer Vision and Pattern Recognition (CVPR’11): June 21-23, 2011. Colorado Springs, USA, 2011: 1297-1304.

[26] Liu Zicheng, Human Activity Recognition with 2D and 3D Cameras[J]. Progress in Pattern Recognition, Image Analysis, and Applications, 2012, 7441: 37

[27] NI Bingbing, WANG Gang, MOULIN P. RGBDHuDaAct: A Colour-Depth Video Database for Human Daily Activity Recognition[C]// Proceedings of IEEE International Conference on Computer Vision Workshops (ICCV Workshops): November 6-13, 2011. Barcelona, Spain, 2011: 1147-1153

[28] LI Wanqing, ZHANG Zhengyou, LIU Zicheng. Action Recognition Based on a Bag of 3D Points[C]// Proceedings of IEEE International Conference on Computer Vision Workshops (CVPR Workshops): June 13-18, 2010. San Francisco, USA, 2010

[29] WANG Jiang, LIU Zicheng, CHOROWSKI J, et al. Robust 3D Action Recognition with Random Occupancy Patterns[C]// Proceedings of the 12th European Conference on Computer Vision — Volume Part II (ECCV’12): October 7-13, 2012. Florence, Italy, 2012: 872-885

[30] Z. Yang, L. Zicheng and C. Hong, "RGB-Depth feature for 3D human activity recognition," in *China Communications*, vol. 10, no. 7, pp. 93-103, July 2013.

[31] W. H. O. Ageing and L. C. Unit, WHO global report on falls prevention in older age: World Health Organization, 2008.

[32] K. I. Withanage, I. Lee, R. Brinkworth, S. Mackintosh and D. Thewlis, "Fall Recovery Subactivity Recognition With

RGB-D Cameras," in *IEEE Transactions on Industrial Informatics*, vol. 12, no. 6, pp. 2312-2320, Dec. 2016.

[33] X. Yang and Y. Tian, "Super Normal Vector for Human Activity Recognition with Depth Cameras," in *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 39, no. 5, pp. 1028-1039, May 1 2017.

[34] Y. Hbali, S. Hbali, L. Ballihi and M. Sadgal, "Skeleton-based human activity recognition for elderly monitoring systems," in *IET Computer Vision*, vol. 12, no. 1, pp. 16-26, 2 2018.

[35] A. Manzi, L. Fiorini, R. Limosani, P. Dario and F. Cavallo, "Two-person activity recognition using skeleton data," in *IET Computer Vision*, vol. 12, no. 1, pp. 27-35, 2 2018.

[36] O. P. Popoola and K. Wang, "Video-Based Abnormal Human Behavior Recognition—A Review," in *IEEE Transactions on Systems, Man, and Cybernetics, Part C (Applications and Reviews)*, vol. 42, no. 6, pp. 865-878, Nov. 2012.

[37] O. Mendels, H. Stern and S. Berman, "User Identification for Home Entertainment Based on Free-Air Hand Motion Signatures," in *IEEE Transactions on Systems, Man, and Cybernetics: Systems*, vol. 44, no. 11, pp. 1461-1473, Nov. 2014

[38] G. Luo, S. Yang, G. Tian, C. Yuan, W. Hu, S. J. Maybank, Learning human actions by combining global dynamics and local appearance, IEEE Transactions on Pattern Analysis and Machine Intelligence 36 (12) (2014) 2466–2482.

[39] Y. Shan, Z. Zhang, P. Yang, K. Huang, Adaptive slice representation for human action classification, IEEE Transactions on Circuits and Systems for Video Technology 25 (10) (2015) 1624–1636.

[40] W. Guo, G. Chen, Human action recognition via multi-task learning base on spatial-temporal feature, Information Sciences 320 (2015) 418–428.

[41] J. Zheng, Z. Jiang, R. Chellappa, Cross-view action recognition via transferable dictionary learning, IEEE Transactions on Image Processing 25 (6) (2016) 2542–2556.

[42] C. Yuan, B. Wu, X. Li, W. Hu, S. J. Maybank, F. Wang, Fusing r features and local features with context-aware kernels for action recognition, International Journal of Computer Vision 118 (2) (2016) 151–171.

[43] K. Soomro, A. R. Zamir, M. Shah, Ucf101: A dataset of 101 human actions classes from videos in the wild, arXiv preprint arXiv:1212.0402, 2012

[44] H. Kuehne, H. Jhuang, E. Garrote, T. Poggio, T. Serre, Hmdb: a large video database for human motion recognition, in: Proceedings of the IEEE International Conference on Computer Vision, IEEE, 2011, pp. 2556–2563

[45] A.-A. Liu, N. Xu, W.-Z. Nie, Y.-T. Su, Y. Wong, M. Kankanhalli, Benchmarking a multimodal and multiview and interactive dataset for human action recognition, IEEE Transactions on Cybernetics, 2017

[46] L. Liu, L. Shao, X. Li, K. Lu, Learning spatio-temporal representations for action recognition: A genetic programming approach, IEEE transactions on cybernetics 46 (1) (2016), 158–170.

[47] A.-A. Liu, Y.-T. Su, W.-Z. Nie, M. Kankanhalli, Hierarchical clustering multi-task learning for joint human action grouping and recognition, IEEE transactions on pattern analysis and machine intelligence 39 (1) (2017) 102–114.

[48] J. Wu, Y. Zhang, W. Lin, Good practices for learning to recognize actions using fv and vlad, IEEE transactions on cybernetics 46 (12) (2016) 2978–2990.

[49] X. Peng, L. Wang, X. Wang, Y. Qiao, Bag of visual words and fusion methods for action recognition: Comprehensive study and good practice, Computer Vision and Image nderstanding 150 (2016) 109–125.

[50] J. Shotton, A. Fitzgibbon, M. Cook, T. Sharp, M. Finocchio, R. Moore, A. Kipman, A. Blake Real-time human pose recognition in parts from single depth images, in: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, IEEE, 2011, pp. 1297–1304.

[51] J. Shotton, R. Girshick, A. Fitzgibbon, T. Sharp, M. Cook, M. Finocchio, R. Moore, P. Kohli, A. Criminisi, A. Kipman, Efficient human pose estimation from single depth images, IEEE Transactions on Pattern Analysis and Machine Intelligence 35 (12) (2012) 2821–2840

[52] L. Xia, C.-C. Chen, J. Aggarwal, View invariant human action recognition using histograms of 3d joints, in: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops, IEEE, 2012, pp. 20–27.

[53] X. Yang, Y. Tian, Super normal vector for activity recognition using depth sequences, in: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2014, pp. 804–811.

[54] H. Rahmani, A. Mahmood, D. Q. Huynh, A. Mian, Hopc: Histogram of oriented principal components of 3d pointclouds for action recognition, in: Computer Vision-ECCV 2014, Springer, 2014, pp. 742–757.

[55] X. Yang, Y. Tian, Eigenjoints-based action recognition using naive-bayes-nearest-neighbor, in: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops, IEEE, 2012, pp. 14–19

[56] J. Wang, Z. Liu, Y. Wu, J. Yuan, Mining actionlet ensemble for action recognition with depth cameras, in: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2012, pp. 1290–1297

[57] F. Ofli, R. Chaudhry, G. Kurillo, R. Vidal, R. Bajcsy, Sequence of the most informative joints (smij): A new representation for human skeletal action recognition, Journal of Visual Communication and Image Representation 25 (1) (2014) 24–38.

[58] W. Li, Z. Zhang, Z. Liu, Action recognition based on a bag of 3d points, in: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops, IEEE, 2010, pp. 9–14.

[59] Zanfir, M., Leordeanu, M., Sminchisescu, C.: 'The moving pose: an efficient 3d kinematics descriptor for low-latency

action recognition and detection'. Proc. of the IEEE Int. Conf. Computer Vision, 2013, pp. 2752–2759.

[60] Ben Tamou, A., Ballihi, L., Aboutajdine, D.: 'Automatic learning of articulated skeletons based on mean of 3d joints for efficient action recognition', *Int. J. Pattern Recogn. Artif. Intell.*, 2017, 31, (04), p. 1750008

[61] Keceli, A.S., Can, A.B.: 'Recognition of basic human actions using depth information', *Int. J. Pattern Recogn. Artif. Intell.*, 2014, 28, (02), p. 1450004

[62] Cal, X., Zhou, W., Wu, L., *et al.*: 'Effective active skeleton representation for low latency human action recognition', *IEEE Trans. Multimed.*, 2016, 18, (2), pp. 141–154.

[63] Hussein, M.E., Torki, M., Gowayyed, M.A., *et al.*: 'Human action recognition using a temporal hierarchy of covariance descriptors on 3d joint locations', *IJCAI*, 2013, 13, pp. 2466–2472.

[64] Xia, L., Aggarwal, J.: 'Spatio-temporal depth cuboid similarity feature for activity recognition using depth camera'. Proc. of the IEEE Conf. Computer Vision and Pattern Recognition, 2013, pp. 2834–2841.

[65] Vieira, A., Nascimento, E., Oliveira, G., *et al.*: 'Stop: space-time occupancy patterns for 3d action recognition from depth map sequences'. Progress in Pattern Recognition, Image Analysis, Computer Vision, and Applications, 2012, pp. 252–259.

[66] Oreifej, O., Liu, Z.: 'Hon4d: histogram of oriented 4d normals for activity recognition from depth sequences'. Proc. of the IEEE Conf. Computer Vision and Pattern Recognition, 2013, pp. 716–723.

[67] Yang, X., Zhang, C., Tian, Y.: 'Recognizing actions using depth motion mapsbased histograms of oriented gradients'. Proc. of the 20th ACM Int. Conf. Multimedia, 2012, pp. 1057–1060.

[68] Zhao, Y., Liu, Z., Yang, L., *et al.*: 'Combing rgb and depth map features for human activity recognition'. Signal & Information Processing Association Annual Summit and Conf. (APSIPA ASC), 2012 Asia-Pacific, 2012, pp. 1–4.

[69] Amor, B.B., Su, J., Srivastava, A.: 'Action recognition using rate-invariant analysis of skeletal shape trajectories', *IEEE Trans. Pattern Anal. Mach. Intell.*, 2016, 38, (1), pp. 1–13.

[70] Yu, G., Liu, Z., Yuan, J.: 'Discriminative orderlet mining for real-time recognition of human-object interaction'. Asian Conf. Computer Vision, 2014, pp. 50–65.

[71] Chaaraoui, A., Padilla-Lopez, J., Flórez-Revuelta, F.: 'Fusion of skeletal and silhouette-based features for human action recognition with rgb-d devices'. Proc. of the IEEE Int. Conf. Computer Vision Workshops, 2013, pp. 91–97.