International Conference on Communication and Signal Processing, July 28 - 30, 2020, India

A Survey on Human Activity Recognition and Classification

Abhay Gupta, Kuldeep Gupta, Kshama Gupta and Kapil Gupta

***Abstract*—Activity Recognition and Classification is one of the most significant issues in the computer vision field. Identifying and recognizing actions or activities that are performed by a person is a primary key goal of intelligent video systems. Human activity is used in a variety of application areas, from human-computer interaction to surveillance, security, and health monitoring systems. Despite ongoing efforts in the field, activity recognition is still a difficult task in an unrestricted environment and faces many challenges. In this paper, we are focusing on some recent research papers on various methods of activity recognition. The work includes three popular methods of recognizing activity, namely vision-based (using pose estimation), wearable devices, and smartphone sensors. We will also discuss some pros and cons of the above technologies and take a view on a brief comparison between their accuracy. The findings will also show how the vision-based approach is becoming a popular approach for HAR research these days.**

***Index Terms*—Deep learning, human activity recognition (HAR), pose estimation, smartphone sensors, wearable sensors.**

I. Introduction

ACTIVITY refers to the movement of the entire body or the different positions of the limbs relative to time against gravity. Human Activity Recognition (HAR) becomes a very popular and active research area for researchers from the last two decades. However, it still remains a complex task due to some unresolvable issues such as sensor movement, sensor placement, background clustering, and the inherent variability of how different people perform activities. Determining detailed activities is beneficial in many areas of human-centric applications, such as home care support, postoperative trauma rehabilitation, abnormal activities, gesture detection, exercise, and fitness. Most of the person's daily tasks can be simplified or automated if recognized by the HAR system. Usually, HAR systems are based on either unsupervised or supervised learning. A supervised system requires pre-training using special datasets, but unsupervised systems have a set of rules during development [1].

Abhay Gupta, Kuldeep Gupta, Kshama Gupta, Kapil Gupta are with Department of Computer Applications, National Institute of Technology, Kurukshetra, India. (phone: +918979578203 e-mail: abhaygupta190@gmail.com kdgupta545@gmail.com kshama30696gupta@gmail.com kapil@nitkkr.ac.in).

This review paper surveys extensively the current progress made towards various activity recognition methods. Moreover, we provide a comprehensive analysis of three techniques, namely wearable devices-based approach, pose based approach, and smartphone sensors. The first one uses sensing devices to be mounted on the subject to collect data from the sensors [2] whereas smartphone sensor based approach takes input from the smartphone sensors such as gyroscope and accelerometer [3] and in the last one activity is classified using pose estimation which require the estimation of body key points through neural network [4]. HAR system can be subdivided into three levels, as follows in Fig. 1.

The rest of the paper is structured as: Section II contains the related work of recent research papers in the field. Section III briefly describes the various methods and techniques used in human activity recognition. A comparative study of different methods and their accuracy is given in Section IV; Section V contains the discussion and our findings; Some concluding remarks and future scope for further discussions are given in Section VI.

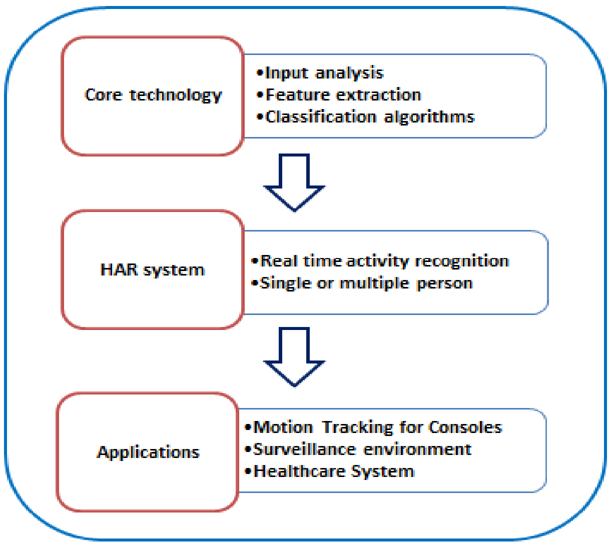

Fig. 1. The overview of general HAR system. This is basic building blocks of almost every system.

978-1-7281-4988-2/20/$31.00 ©2020 IEEE

II. RELATED WORK

In recent years, recognition and understanding of human behaviour have received much attention. Different techniques have been used to understand patterns of behaviour activity and to understand the scene. In this work, we have reviewed some selected papers from 2016-2019 on HAR. Our study provides a brief analysis of HAR methods. Ghazal et al. [5] proposed a pose based HAR, it uses openpose library and forward feed CNN to predict a confidence map of 18 key points and further uses the decision-making algorithm to classify standing and sitting activity. Tsai et al. [6] presented a system for detecting the activities using a Kinect sensor, and 11 kinds of activities are recognized through discrete Hidden Markov models (HMMs). The system is developed for training the robot. Gatt et al. [7] proposed an approach for detecting abnormal behaviour such as fall detection. The work uses PoseNet and Openpose pre-trained pose estimation model, then Long Short-Term Memory (LSTM) and CNN are used for activity classification.

Bulbul et al. [3] recognize human activity by using different classification and machine learning approaches such as Bagging, k-NN, etc. For this, they have used two smartphone sensors, accelerometers, and gyroscope and recognize 6 different activities. Tran et al. [8] proposed a three smartphone sensors-based approach for HAR. they used SVM for the classification and identification of activity and optimized the classification model to identify the activities. RoyChaudhury et al. [9] used a single smartphone sensor for HAR. they use different classifiers to test the proposed model like trees (Complex tree). They considered 12 activities, including static and dynamic activities, for his work.

III. VARIOUS APPROACHES TO HAR

To achieve the goal of recognizing human activity, a HAR system is required. The two most commonly used techniques for this purpose are sensor-based and vision-based activity recognition. We can classify them, as shown in Fig. 2.

A. Pose Based Approach

Poses are important for analyzing videos, which include humans, and there is strong evidence that body posture concepts are very effective for various tasks such as activity recognition, content extraction, etc. This approach classifies human actions based on the coordinate information of the body parts. Basically, HPE refers to the process of assessing the composition of a part of the human body (3D poses) or the projection onto an image plane (2D HPE). It covers all issues related to the human body, from understanding the entire human body to the detailed localization of body parts [4]. It is formulated as a regression problem that can be modelled with a simple CNN. It takes the entire image as input and shows the pixel coordinates of the body's key points. There are 15 body joints: Neck, LKnee, LAnkle, RShoulder, RWrist, Relbow, LShoulder, RHip, LElbow, LWrist, Chest, and 14 joint connections. The classification problem can be formulated as a multi-class classification problem that can also be modeled by using neural networks. CNN accepts the body joints location as input and generates a number vector representing the probability of each activity labels accordingly. Some popular dataset is also available such as MPII, which contain more than 20,000 labelled images of 410 specific subcategories activities under 20 activity categories [10].

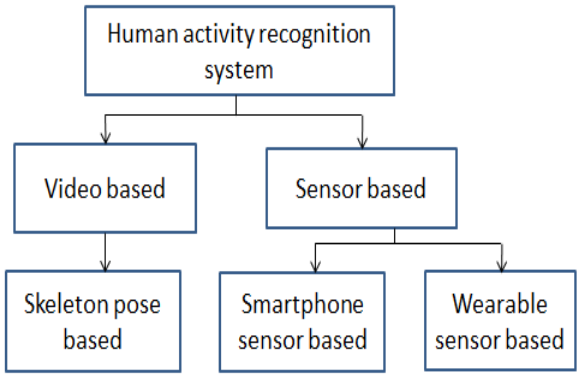

Fig. 2. Classification of HAR system based on their approaches.

B. Smartphone Sensor- Based

Smartphones are the most useful tool in our everyday lives, and advanced technology is enabling us to meet the needs and expectations of customers every day. To make these devices more functional and powerful, designers are adding new modules and devices to the hardware. Sensors enhance the capabilities of smartphones and play a major role in understanding the environment. As a result, most smartphones have a variety of built-in sensors that can collect a wealth of useful data about the human's daily life.

Sensors retrieve information from body gestures and then recognized the activities. The most commonly used sensors are accelerometer and gyroscope etc. Accelerometer sensor is used for measures the change in speed, and gyroscope is used to measures the orientation of the body. Some techniques of HAR through smartphones used in recent studies are SVM, k-NN, Bagging, Ada Boost [3].

C. Wearable Sensor Based

The wearable technique uses sensing devices to be mounted on the subject to collect data from the sensors. As human activity contains actions of different bodily positions, the research of human activity needs to capture information from more than one sensor installed on the different parts of the body of the person. Wearable devices must be designed with user accessibility in mind. Lightweight, modern, and comfortable wearing devices with embedded sensors are used for activity monitoring.

Activity monitoring sensors are used in multiple datasets. The most commonly used sensors are an accelerometer, gyroscope, magnetometer, and RFID tag [11]. After feature

TABLE I
COMPARATIVE ANALYSIS OF HUMAN ACTIVITY RECOGNITION ALGORITHM BASED ON PREVIOUS STUDIES

S.No.	AUTHORS AND YEAR	FEATURES	REFINE FEATURES	DATASET	ACTIVITIES	MODEL USED	ACCURACY (%)
1.	Bulbul et al., 2018 [3]	Smartphone sensor (accelerometers and gyroscope) to recognize human activities	Noise is filtered using median and 20Hz Butterworth filter and then 3Hz Butterworth filter to filter the result	Dataset collected from 9 individuals (2947 records with 561 features)	Recognize 6 different activities like walking, sitting, climbing up the stairs, etc.	SVM	99.4
						k-NN	97.1
						AdaBoost	97.4
						Bagging	98.1
2.	Tran et al., 2016 [8]	Three smartphone sensors (accelerometers, gyroscope, and accelerometer sensor linearity) used to recognize human activities	Not available	Data collected from 10 volunteers contains 10,939 samples	Recognize 6 different activities like walking, sitting, upstairs, etc.	Support Vector Machine (SVM)	89.59
3.	RowChowdhury et al., 2018 [9]	Smartphone sensor (accelerometer) used to recognize human activities	Median filter and Butterworth filter to removes the low-frequency acceleration (gravity) and noise from accelerometer signal.	They have collected their own dataset	Recognize 12 different activities like Sit on floor, Climbing upstairs, etc.	SVM	89.5
						k-NN (k=1)	90.9
						Ensemble (Bagged Trees)	94.2
						Complex Tree	91.7
4.	Nandy et al., 2019 0	Smartphone inertial accelerometer and heart rate sensor to recognize human activities.	Butterworth filter used for to filter noise	Collected their own dataset	Recognize 6 different activities like sitting, walking, running, lying down, standing, climbing stairs, etc.	Linear Regression	53.92
						Decision Tree	93.54
						Multilayer Perceptron	77.07
						Gaussian Nave Bayes	73.73
5.	Khokhlov et al., 2018 [2]	Accelerometer and gyroscope sensor mounted on the body	Not available	Collected own dataset	Recognize three activities, such as sitting, walking, standing, etc.	J48	98.4
						Naïve Bayes	76.2
						Random Forest	95.0
						KNN (k=1)	93.6
6.	Ghazal et al., 2018 [5]	18 body key points obtained	Openpose library used to extract features	Random images (146) downloaded from the internet	Recognize activities such as Sitting on the ground or chair, and standing.	Feedforward CNN and decision-making algorithm	95.2
7.	Tsai et al., 2017 [6]	Vector Quantization is used to reduce the noise through clustering.	Not available	Images for the dataset from Kinect sensor (275 samples)	11 different kinds of activities for training robots.	Vector Quantization - Hidden Markov Model (VQ-HMMs)	95.64
8.	Gatt et al., 2019 [7]	PoseNet and Openpose pre-trained models used	Not available	COCO dataset	Recognize abnormal activity such as fall detection	Semi-supervised LSTM and CNN for classification	93

extraction and modelling, human activities can be recognized through statistics, and machine learning algorithm is applied. How to map low-level sensor data to higher-level abstractions is the key to activity recognition.

IV. Comparative Study

In this section, a comparative analysis of various activity recognition system is taken out based on the literature study we have done so far. In Table I. we have shown accuracies of different machine learning algorithms used in the developing HAR system.

V. Findings and Discussion

Fig. 3. shows the graph of techniques used in HAR in recent years. Our review results depict that both smartphone and wearable sensor technology are common in HAR research. On the other hand, the pose-based approach is not so popular in the early days, probably due to the limitation of scenes and human movements in 3D space. In addition, detecting and extracting people from image sequences is another limitation that requires sophisticated machining. Therefore, when large volumes of data are processed at once, real-time HAR system can achieve better results. The pose is estimated by cameras so, another concern raised by these HAR systems is privacy issues. A person can be uncomfortable or forced to look at the watch at all times. Though it is useful to recognize any activities with video cameras, most of all, they need infrastructure support as it needs to install cameras in surveillance places and is highly dependent on illumination [9].

Talking about other technologies, smartphones are easy to carry and are used in everyday life, so it gives an edge to use them for activity monitoring. Although there are some practical limitations of working with real-time data from the app, such as the type of smartphone used, sensor frequency, smartphone location, etc. On the other side, wearable devices are becoming popular because they are easier to implement, available at low cost, operate in harsh environments, and also do not interfere with privacy. Despite having many benefits of wearable sensors still, it has some drawbacks. HAR carrier systems typically require multiple sensors to be worn or mounted on different parts of the body, which is troublesome, disturbing, and uncomfortable for the subject [11].

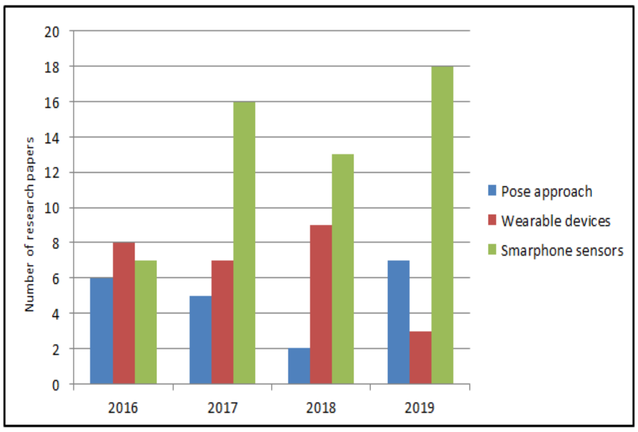

Fig. 3. HAR papers published in recent years.

In terms of universal context, there is no specific indicator or measurement that indicates whether wearable sensors are better than smartphone sensors or vice versa. Both sensors are expected to have advantages and disadvantages depending on the subject and the purpose of use. Therefore, researchers and developers need to figure out the subjects and their use before using any HAR technology [12-14].

VI. Conclusion and Future Scope

A review has been completed on some selected research papers published in 2016-2019 on various HAR technologies. We have categorized these technologies into three main categories, namely HAR using pose estimation (vision-based), smartphone sensors, and wearable sensors. From our study, it is found that the emergence of wearable technology has become a better solution for providing support services to people. However, the system still has some limitations. Some actions have low recognition rates. Further research is needed to improve accuracy and increase the number of activities detected by the system. Also, we noticed that vision-based approach is not much popular among the three in last two decades in spite of having the better results due to its limitations but in the upcoming years with the advancement in the technology machines with high computational power is made available easily, which are capable of processing large amount of data within less time vision-based approach will become a great choice for HAR.

References

[1] O. C. Ann and L. B. Theng, "Human activity recognition: A review," 2014 IEEE International Conference on Control System, Computing and Engineering (ICCSCE 2014), Batu Ferringhi, 2014, pp. 389-393.

[2] I. Khokhlov, L. Reznik, J. Cappos and R. Bhaskar, "Design of activity recognition systems with wearable sensors," 2018 IEEE Sensors Applications Symposium (SAS), Seoul, 2018, pp. 1-6.

[3] E. Bulbul, A. Cetin and I. A. Dogru, "Human Activity Recognition Using Smartphones," 2018 2nd International Symposium on Multidisciplinary Studies and Innovative Technologies (ISMSIT), Ankara, 2018, pp. 1-6

[4] Bearman, Amy L., Stanford and Catherine Dong. “Human Pose Estimation and Activity Classification Using Convolutional Neural Networks.” (2015).

[5] S. Ghazal and U. S. Khan, "Human posture classification using skeleton information," 2018 International Conference on Computing, Mathematics and Engineering Technologies (iCoMET), Sukkur, 2018, pp. 1-4.

[6] A. Tsai, Y. Ou, C. Sun and J. Wang, "VQ-HMM classifier for human activity recognition based on R-GBD sensor," 2017 International Conference on Orange Technologies (ICOT), Singapore, 2017, pp. 201-204.

[7] T. Gatt, D. Seychell and A. Dingli, "Detecting human abnormal behaviour through a video generated model," 2019 11th International Symposium on Image and Signal Processing and Analysis (ISPA), Dubrovnik, Croatia, 2019, pp. 264-270.

[8] D. N. Tran and D. D. Phan, "Human Activities Recognition in Android Smartphone Using Support Vector Machine," 2016 7th International

Conference on Intelligent Systems, Modelling and Simulation (ISMS), Bangkok, 2016, pp. 64-68

[9] I. RoyChowdhury, J. Saha and C. Chowdhury, "Detailed Activity Recognition with Smartphones," 2018 Fifth International Conference on Emerging Applications of Information Technology (EAIT), Kolkata, 2018, pp. 1-4

[10] A. Singh, S. Agarwal, P. Nagrath, A. Saxena and N. Thakur, "Human Pose Estimation Using Convolutional Neural Networks," 2019 Amity International Conference on Artificial Intelligence (AICAI), Dubai, United Arab Emirates, 2019, pp. 946-952.

[11] R. Liu, T. Chen and L. Huang, "Research on human activity recognition based on active learning," 2010 International Conference on Machine Learning and Cybernetics, Qingdao, 2010, pp. 285-290.

A. Nandy, J. Saha, C. Chowdhury and K. P. D. Singh, "Detailed Human Activity Recognition using Wearable Sensor and Smartphones," 2019 International Conference on Opto-Electronics and Applied Optics (Optronix), Kolkata, India, 2019, pp. 1-6.

[12] S. N. Boualia and N. Essoukri Ben Amara, "Pose-based Human Activity Recognition: a review," 2019 15th International Wireless Communications & Mobile Computing Conference (IWCMC), Tangier, Morocco, 2019, pp. 1468-1475.

[13] A. Das Antar, M. Ahmed and M. A. R. Ahad, "Challenges in Sensor-based Human Activity Recognition and a Comparative Analysis of Benchmark Datasets: A Review," 2019 Joint 8th International Conference on Informatics, Electronics & Vision (ICIEV) and 2019 3rd International Conference on Imaging, Vision & Pattern Recognition (icIVPR), Spokane, WA, USA, 2019, pp. 134-139.

[14] M. Ahmed, A. Das Antar and M. A. R. Ahad, "An Approach to Classify Human Activities in Real-time from Smartphone Sensor Data," 2019 Joint 8th International Conference on Informatics, Electronics & Vision (ICIEV) and 2019 3rd International Conference on Imaging, Vision & Pattern Recognition (icIVPR), Spokane, WA, USA, 2019, pp. 140-145.