A Review of State of Art Techniques for 3D Human Activity Recognition System

Bhavana Sharma and Jeebananda Panda

Abstract Recognizing human activities through video sequences and images is still a challenge due to background jumble, partial occlusion, changes in scale, viewpoint, lighting and appearance. A human activity classification technique has been comprehensively reviewed by the researchers. We have categorized human activity methodologies with object detection and feature extraction along with their subcategorization, advantages and restrictions. Moreover, we provide a comprehensive analysis of the existing, publicly available human activity datasets with applications and examine the prerequisites for an ideal human activity recognition dataset. At last, we present some open issues on human activity recognition and characteristics of future research directions.

Keywords Human activity recognition · Object detection · Feature extraction · Object classification · HAR datasets

1 Introduction

Population of elders is increasing with a rapid rate in most of the western countries and hence the challenges [1]. If we convert this in the form of percentage, by 2050, it will be 30% in Europe and China which is maximum globally and then 20.2% in United States of America (USA). This fact was established after a survey of WHO, that on an average, 28–35% of elderly people meet with an accident, because of falling, annually. According to WHO report, 37 million fall accidents are reported every year out of which 64.6 thousand people lose their life because of these accidents [2, 3]. In today's modern world of nuclear families, elderly people are living alone at their homes and they are more prone to meeting with such accidents while staying

B. Sharma (✉) · J. Panda
Department of Electronics and Communication Engineering, Delhi Technological University, New Delhi, India
e-mail: bhavanasharma.iec@gmail.com

J. Panda
e-mail: jpanda@dce.ac.in

R. Agrawal et al. (eds.), *Modern Electronics Devices and Communication Systems*, Lecture Notes in Electrical Engineering 948,
https://doi.org/10.1007/978-981-19-6383-4_1

at home. Hence, the study of fall detection, to improvise the science of detection, is more crucial and important [4–6]. In human activity recognition, to solve the existing and upcoming challenges toward activity recognition, researchers are using different techniques to beat the challenges of analysis. Based on the defined categories in different areas, methods and approaches may differ from one to another. Most commonly used cases of activity recognition is in medical and surveillance, where we talk about the devices and systems which are beneficial to humankind, in improving the life by mitigating the threats. By having a direct impact on saving lives, researchers are very keen to work in the field of video surveillance.

Human activity recognition through a comprehensive survey covers human activity recognition, i.e., 2D and 3D HAR based on RGB, depth and skeleton-based methodologies. The literature is updated with the application of recent advances field of human action recognition in Sect. 2. A structured arrangement of 2D and 3D object detection techniques has been discussed in Tables 1 and 2 highlighting different feature extraction techniques. Organization of our survey is as follows. Section 2 provides a panoramic summary of the related state-of-the-art survey works in the area of abnormal human action recognition followed by paper count analysis per year. It will help the reader to get an overview of key contributions of previous surveys done. Section 3 provides that human action recognition system is closely discussed with methodologies for detection, extractions and classification techniques evolved. Sections 4 and 5 outline recently introduced publicly available datasets used for activity recognition with challenges and applications. Finally, peculiar observations and possible directions are highlighted that need to further explore for research in the field of HAR.

2 Literature Survey

Shian-Ru Ke presented trends of HAR in video signals, and article explains the three different areas in activity recognition using core technology, human activity recognition systems and applications. This article throws light on application areas like surveillance, healthcare and entertainment industry where major focus is on surveillance in healthcare including its challenges [7]. Pau Climent-Pérez's article is based on HBA for ambient assisted living using AI. This study beautifully covers the estimation based on pose and gaze for movement identification. Later, it represents the latest work showcasing latest data tools and new datasets are described here [8]. Paul explained the techniques which are used in identification of human objects in surveillance video data with a benchmark datasets including directions for further research in living human identification and detection [9]. Fei Han represented an extensive survey of space time representations of human based on 3D skeletal data on categorization and analysis including modality, feature engineering, structure and transition including representation encoding [10]. Tej Singh explained key specifications of vision-based human activity recognition datasets which are discussed along with the algorithms according to the datasets best performance. Resolutions, actions/actors,

Table 1 Comparison of object detection methods

Techniques		Accuracy	Computational time	Advantages	Disadvantages
Background subtraction	Mixture of Gaussian model	Moderate	Moderate	Better response with Simple implementation and multi-modal scenarios	Not suitable for dynamic background and need to defined parameters
	Non-parametric background model	Moderate to high	Low to moderate	With significant post-processing, performs better in moving background	In occlusion, cannot performed
	Temporal differencing	High	Low to moderate	With sudden illumination changes, gives well performance in indoor environment	
	Warping background	High	Moderate to high	With high dynamic background, it is good in outdoor environment	Cannot work with occlusion
	Hierarchical background model	High	Low to moderate	Block-based and pixel-based approaches both are used and faster than pixel-based approach	Not good quality
Optical flow		Moderate	High	Good with dynamic camera and crowd detection	Highly computation intensive
Spatio-temporal filter		Moderate to high	Low to moderate	Perform good with low-resolution scenarios	More noise

frame rate, background and application domain are discussed in the paper [11]. Allah Bux described the image segmentation techniques and reviewed including challenges and future scope of research [12]. Athanasios Lentzas focused on the ABHAR for senior citizens. Analysis is done based on the taxonomy [13]. Michalis Vrigkas provided a comprehensive analysis of available datasets and examine the

Table 2 Comparison between different feature extraction methods

Techniques	Accuracy	Computational time	Advantages	Disadvantages
Shape-based method	Moderate	Low	With appropriate templates a simple pattern-matching approach is used	Not able to determine internal movements and in dynamic situations cannot performed
Motion-based method	Moderate	High	There is no need to predefined pattern templates	Cannot identify a non-moving human
Texture-based method	High	High	Good quality	More computation time

requirements for ideal datasets [14]. Tej Singh focused on the issue of benchmark datasets. Here article provide the comprehensive review to address this issue of benchmark datasets, action recognition-related RGB-D video datasets with 27 single-view datasets, 10 multi-view datasets, are provided [15], and various human activity recognition handcrafted and deep approaches are explained with 2D and 3D RGB and RGB_D dataset in this paper [16].

3 Methodology

See Fig. 1.

3.1 *Input Image/Video*

In the general process of recording the procedure of RGB cameras, a differentiation in the activity was created to analyze the sequence of actions, but at the same time, we have the challenges related to background clarity and lightning effect of the images which leads to the complexity while working on a design of the solution [17, 18]. Afterward, there was a regular improvement in the research methods to improve the factors like capturing of the depth of action in an optimal cost and real-time with the help of infrared radiation which provided a relief to the challenges related to lightning effects.

Fig. 1 Process of human activity detection

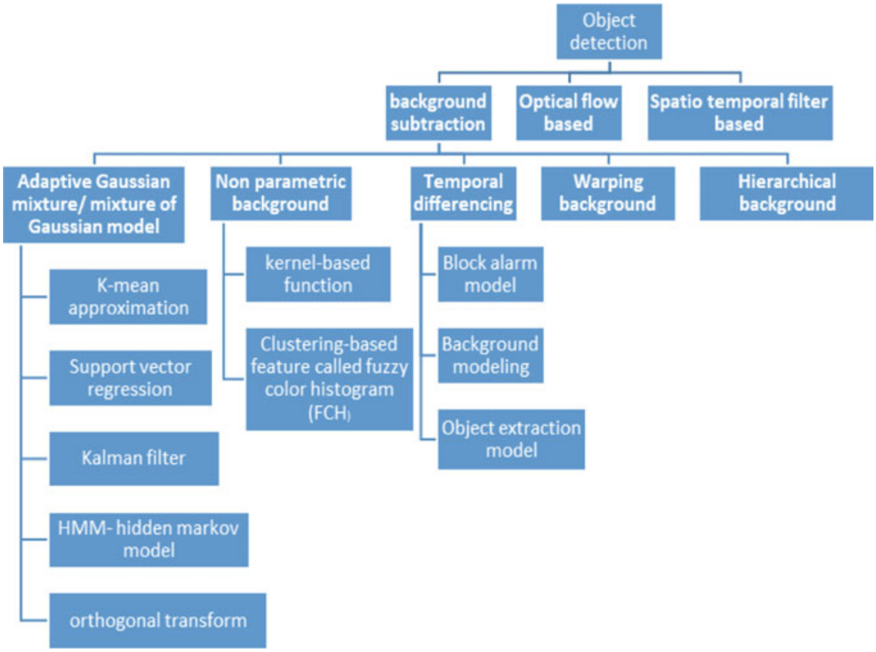

Fig. 2 Three types of human object detection

3.2 Object Detection

Background subtraction—In this techniques, a comparison of the moving object has been done based on the difference between current frames with the background frame. This comparison is done either pixel by pixel or block to block [19–21].

Optical flow based—Here in this technique, detection of moving object w.r.t. time based on the characteristic of flow vector has been used. There are challenges also related to lightening effect, motion sensitivity or noise which leads to high computational time [22].

Spatio-temporal filter based—This method is used to have reduced the computational requirement and noise by using data volume spanned by the moving object in a video signal [23]. This method is called 3D spatio-temporal because it works with spatial as well as time (Fig. 2).

3.3 Feature Extraction

This is a technique for the reduction of data dimension in which transforms lower dimension data into a modified featured space. Researcher selects a subset of features from the superset which will meet the forecasting requirements of target labels

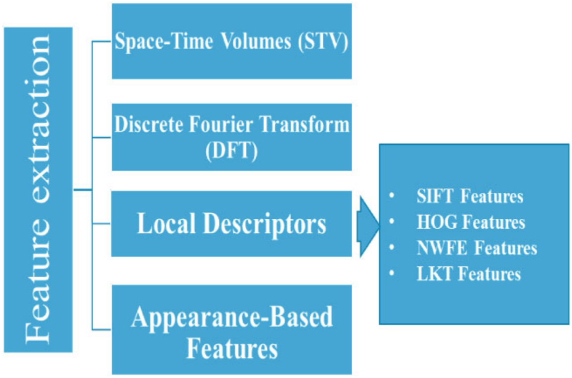

Fig. 3 Types of feature extraction

correctly diminishing the complexity of computation of different algorithm of learnings and predictions by subtracting the cost of remaining features left in the list [24]. Out of all methods, principal component analysis (PCA) gives the most reliable results in reduction of dimensions and extraction of attributes in case of linear structures [25, 26]. On the other hand, for nonlinear structure, linear discriminant analysis (LDA) technique is used to mitigate the challenges of PCA [27]. Linear discriminant analysis is used to separate the features with the aim of establishment of a linear transformation to attain the biggest class discrimination. The traditional LDA is used to find out a standard discriminant subspace (Fig. 3).

3.4 Object Classification

Shape-based method—The shape data of moving regions such as points, boxes and blobs is described foremost and after that deemed as a standard pattern recognition. While using the aforementioned approach, the large number of possible impressions of the body creates chaos between a moving human and other moving objects [28, 29]. An enormous challenge with this method is that it cannot apprehend the internal motion of the object in the contour area.

Motion-based method—In this method, we can overcome with the confusion between a moving human and other moving objects by using object motion characteristics with patterns analysis means to identify people in other moving objects, it uses the periodic property of captured images [30–32].

Texture-based method—Texture-based methods use intensity patterns for nearby pixels. This technique counts the gradient directions of local area of image and does calculations on a dense grid of evenly spaced cells. For better accuracy, it uses overlapping local contrast normalization.

4 Datasets

There are some important datasets.

KTH human motion dataset—This dataset contains six human actions performed by 25 subjects in four different situations. Running, jogging, walking, boxing, hand waving or clapping are performed in more than 2000 sequences. The backgrounds are homogeneous and uncluttered. Video files are classified by operation, to eliminate unnecessary operations easily [13].

Weizmann human action dataset—It uses static front-side cameras to record individual human movements from 10 subjects in different environments. Approximately 340 MB of video sequences are available. The actions performed include bending, walking, running, hand waving and different types of jumping.

HOHA—(Hollywood human actions)—HOHA dataset contains video sequences from 32 movies with annotations for eight action types: AnswerPhone, GetOutCar, HandShake, HugPerson, Kiss, SitDown, SitUp and Stand.

INRIA Xmas motion acquisition sequences—The video images of 390×291 pixel which is recorded from five different angles are included in these sequences. 11 actors perform 13 actions: check watch, cross arms, scratch head, sit down, get up, turn around, walk, wave, punch, kick, point, pick up, throw overhead and throw from bottom to top.

TUM kitchen dataset—This dataset aims ADLs in a kitchen scene with a low level of action. Multiple subjects perform tablet setting in different ways; transporting items one by one; and other behaviors are natural, grabbing multiple objects at once.

5 Challenges in HAR Dataset

In this section, we discuss the various current challenges in the dataset.

Background and environment conditions—In videos if there is moving object or background, it is very difficult to recognize human activity. There are so many types of background in a video signal like slow and fast, dynamic and static, airy and rainy, and crowded. Same recognition activity in environment conditions which contains various issues like rain, waves, trees and water is affected.

Similarity and Difference of actions—There are many actions that looks same in the videos like running, jogging, walking, etc. The same type of procedure affects classification accuracy. Similarity between classes of actions in datasets provides a fundamental challenge.

Occlusion—Occlusion occurs when an object is hiding the another object. The occlusion can be classified into two parts one is self-Occlusion and another one is partial occlusion. Occlusion is a greater challenge in computer vision applications such as human posture, object tracking and video monitoring.

View variations—In human identification system, any action recorded inside the video is the most crucial characteristic. Multiple views have larger facts comparetively single view which leads to fair analysis of captured perspective in dataset.

6 Conclusion and Future Work

The literature survey encloses a wide area around HAR covering different methodologies and techniques of identification, detection and limitations along with its pros and cons. It also throws light on dataset benchmark and its quality which leads to the variation in results. Numerous HAR dataset challenges discussed.

In future, HAR systems need to address specific issues related to the quality of dataset and connect it to the real-life application development. In future, researcher will need to work on the challenges relating to noise, input quality data and various process-related challenges. Some meaningful datasets to represent abnormal actions in different scenarios are still a problem. Working on deep architectures from primary CNN to RCNN, RNN, auto-encoders can be extended to enhance the parameters of recognition systems.

References

1. Merrouche F, Baha, N (2016) Depth camera based fall detection using human shape and movement. In: IEEE international Conference on signal and image processing
2. Ma X, Wang H, Xue B, Zhou M, Ji B, Li Y (2014) Depth-based human fall detection via shape features and improved extreme learning machine. IEEE J Biomed Health Inf
3. Bian Z-P, Chau L-P, Magnenat-Thalmann N (2014) Fall detection based on body part tracking using a depth camera. IEEE J Biomed Health Inf
4. Lentzas A, Vrakas D (2019) Non-intrusive human activity recognition and abnormal behavior detection on elderly people: a review. Springer Nature B.V
5. Pham C, Nguyen-Thai S, Tran-Quang H, Tran S, Vu H, Tran T-H, Le T-L (2020) SensCapsNet: deep neural network for non-obtrusive sensing based human activity recognition. IEEE Access
6. Popoola OP, Wang K (2012) Video-based abnormal human behavior recognition—a review. IEEE Trans Syst Man Cybern C: Appl Rev
7. Ke S-R, Thuc HLU, Lee Y-J, Hwang J-N, Yoo J-H, Choi K-H (2013) A review on video-based human activity. Recognition 2:88–131. https://doi.org/10.3390/computers2020088
8. Chaaraoui AA, Climent-Pérez P, Flórez-Revuelta F (2012) A review on vision techniques applied to human behaviour analysis for ambient-assisted living. Elsevier
9. Paul M, Haque SME, Chakraborty S (2013) Human detection in surveillance videos and it applications—a review. EURASIP J Adv Signal Process
10. Han F, Reily B, Hoff W, Zhang H (2016) Space-time representation of people based on 3D skeletal data: a review. Elsevier
11. Dhiman C, Vishwakarma DK (2019) A review of state-of-the-art techniques for abnormal human activity recognition. In: Engineering Applications of Artificial Intelligence Elsevier, pp 21–45
12. Dhiman C, Vishwakarma DK (2020) View-invariant deep architecture for human action recognition using two-stream motion and shape temporal dynamics. IEEE Trans Image Process
13. Singh T, Vishwakarma DK (2019) Human activity recognition in video benchmarks: a survey. Springer Nature Singapore

14. Jankowski S, Szymański Z, Mazurek P, Wagner J (2015) Neural network classifier for fall detection improved by Gram-Schmidt variable selection. In: The 8th IEEE international conference on intelligent data acquisition and advanced computing systems
15. Brun L, Percannella G, Saggese A, Vento M IAPR Fellow (2017) Action recognition by using kernels on aclets sequences. Elsevier
16. Jing C, Wei P, Sun H, Zheng N (2019) Spatiotemporal neural networks for action recognition based on joint Loss. Springer-Verlag, London Ltd., part of Springer Nature
17. Thien Huynh- Cam-Hao Hua, Nguyen Anh Tu , Taeho Hur , Jaehun Bang , Dohyeong Kim , Muhammad Bilal Amin , Byeong Ho Kang , Hyonwoo Seung , Soo-Yong Shin , Eun-Soo Kim , Sungyoung Lee (2018) "Hierarchical topic modeling with pose-transition feature for action recognition using 3D skeleton data", Elsevier
18. Sarakon S, Tamee K (2020) An individual model for human activity recognition using transfer deep learning. In: Joint international conference on digital arts
19. Cai X, Zhou W, Wu L, Luo J, Li H (2016) Effective active skeleton representation for low latency human action recognition. IEEE Trans Multimedia 18(2)
20. Suto J, Oniga S, Lung C, Orha I (2018) Comparison of offline and real-time human activity recognition results using machine learning techniques. Springer
21. Dhiman C, Vishwakarma DK (2019) A robust framework for abnormal human action recognition using R-transform and Zernike moments in depth videos. IEEE Sens J
22. Ladjailia A, Bouchrika I, Merouani HF, Harrati N, Mahfouf Z (2019) Human activity recognition via optical flow: decomposing activities into basic actions. Springer-Verlag London Ltd., part of Springer Nature
23. Ji X, Cheng J, Feng W, Tao D (2017) Skeleton embedded motion body partition for human action recognition using depth sequences. Elsevier
24. Vishwakarma DK, Rawat P, Kapoor R (2015) Human activity recognition using Gabor wavelet transform and Ridgelet transform. In: 3rd international conference on recent trends in computing—ICRTC
25. Lahiri D, Dhiman C, Vishwakarma DK (2017) Abnormal human action recognition using average energy images. In: Conference on information and communication technology
26. Abdull Sukor AS, Zakaria A, Abdul Rahim N (2018) Activity recognition using accelerometer sensor and machine learning classifiers. In: 2018 IEEE 14th international colloquium on signal processing & its applications (CSPA 2018), Penang, Malaysia, 9–10 March [2018]
27. Tao D, Jin L, Yuan Y, Xue Y (2016) Ensemble manifold rank preserving for acceleration-based human activity recognition. IEEE Trans Neur Netw Learn Syst
28. Akagündüz E, Aslan M, Şengür A, Wang H, İnce MC (2015) Silhouette orientation volumes for efficient fall detection in depth videos. IEEE J Biomed Health Inf
29. Mazurek P, Morawski RZ (2015) Application of Naïve Bayes classifier in fall detection systems based on infrared depth sensors. In: The 8th IEEE international conference on intelligent data acquisition and advanced computing systems
30. Wagner J, Morawski RZ (2015) Applicability of mel-cepstrum in a fall detection system based on infrared depth sensors. In: The 8th IEEE international conference on intelligent data acquisition and advanced computing systems
31. Jankowski S, Szymański Z, Dziomin U, Mazurek P, Wagner J (2015) Deep learning classifier for fall detection based on IR distance sensor data. In: The 8th IEEE international conference on intelligent data acquisition and advanced computing systems
32. Zhang H, Parker LE (2011) 4-dimensional local spatio-temporal features for human activity recognition. In: IEEE international conference on intelligent robots and systems, San Francisco