Survey on Human Activity Recognition

Piyush Soni[1], Dr. Sunil Phulre[2], Dr. Sadhana K Mishra[3]

[1]M. Tech. Scholar, Department of CSE, LNCT, Bhopal (India)

[2]Associate Professor, Department of CSE, LNCT, Bhopal (India)

[3]Head & Professor, Department of CSE, LNCT, Bhopal (India)

Abstract: *Recognizing the activities of human beings is viewed as extremely fundamental in person-to- person communication and relational relationships as it delivers information with respect to the individuality of a group, their character and mental condition. Digging out this data is not an easy job. The significant concept of investigation of AI and machine Learning is the capability of a person to distinguishing the exercises of someone else. The human gestures are complex and also very dynamic. This paper discusses thoughts of different researchers.*

Keywords: Human Activity Recognition, Artificial Intelligence, Machine Learning.

Introduction

There is always some motive behind every human activity. To understand the behavior and its interaction of human with environment automatically is become an interesting research field in the previous few years as its application is capable in several domains. There are an enormous number of applications including intelligent video surveillance and to monitor environmental home, to store and retrieve video etc. available in HAR. Various research topics in computer vision are covered while recognizing the human action such as detecting the human in video, estimating the human pose and analyzing and understanding of time series data. The human action must be recognized in person-to-person interaction and interpersonal relations as it offers the information related to the personality of someone, their behavior and mental state. It is quite difficult to extract this information.

II. Literature Survey

Xinyi Liang, et.al (2019) conferred that it was difficult to recognize the action in computer vision [1]. An activity identifying approach for human body was intended here. This approach, employing combined trajectory photographs and visual facilitation, was used for modeling the activities of people. Initially, human skeleton data was extracted from an RGB-cam implementing OpenPose. The activities of individuals were revealed in a video as the trajectory of human skeletal joints within a photograph. The operation of OCR had a great impact of on action recognition. Thus, the issue to recognize action was considered as the issue of trajectory image identification. The action was recognized with the deployment of integration of Histogram of Oriented Gradients with Support Vector machine. Eventually, the conducted

tests analyzed and compared the performance of intended approach. The intended scheme assisted researchers who focused on using the joint trajectory images to recognize the activities of individuals.

Qinkun Xiao, et.al (2017) designed a DNN framework for recognizing the human action. An autoencoder and PRNN were comprised in this framework [2]. At first, an overlay of binary images was set up for employing as training data by securing outlines of human body for every image frame and integrating them under the system learning phase. The training of autoencoder was completed on the basis of DNN for the extraction of action attribute. After that, the Pattern Recognition Neural Network was trained through the supervised learning. At last, the autoencoder was incorporated with PRNN for designing a novel DNN that was known as the APRNN. The designed model had provided the best performance. The translation of human action sequences was done into the binary overlay images and the actions were recognized using the designed model under the action recognition phase. The outcomes obtained in testing demonstrated that the presented framework had performed more efficiently in comparison with the state-of-art approaches.

Thien Huynh-The, et.al (2020) suggested a various leveled 3D skeleton dependant framework developed by fusing deep features to understand human activity [3]. The CNNs were executed to provide the deep information that was employed to model the human appearance and action dynamic. A multi-stream Convolutional Neural Network algorithm assisted in Extracting the deep attributes so that the concealed correlations were exposed in the spatial and temporal dimensions. The NTU RGB+D dataset was deployed for carrying out the experiment. The outcomes of experiments indicated that the recommended model had provided better performance as compared to various DL based approaches of action identification.

B Jagadeesh, et.al (2016) described those human actions were detected and recognized based on video on the KTH dataset and on videos of real-time [4]. Initially, one hundred frames were abstracted from each video clip, and the optical flow was evaluated amid frames. The extracted was transformed into binary image. Subsequently, the feature vector was extracted from the binary images through the HOG descriptor. The extracted feature vectors were employed as training attributes in SVM classification algorithm for preparing a trained model. The real time videos were built in which various human actions such as walking, jogging etc were included to perform testing. Similar type of attributes had extracted and utilized in the Support Vector machine to classify these actions.

Amir Nadeem, et.al (2020) paid attention on integrating LDA with ANN to identify and trace precise activities of individuals [5]. The complicated human actions were detected using 2 existing datasets named KTH-dataset and Weizmann Human Action. There were multidimensional attributes which had estimated using body-models achieved from 12 parts of body. The ANN had employed these attributes as its inputs. The efficacy of presented technique was evaluated by comparing its results with other existing models. The outcomes of experiment revealed the reliability and applicability of presented method in cyberspace, smart picture recovery and man machine communication.

Yun Han, et.al (2018) developed a GSA system to demonstrate the loads for skeleton joints and the thought of ALC model to accumulate the frames which assumed a huge part in decision making [6]. The

suggested global spatial attention was combined with accumulative learning curve in LSTM so that a robust action recognition framework was developed in which the human skeletal joints were considered as input. This model assisted in predicting the human action using the enhanced through the STA. The NTU data sets were utilized to conduct the experiments. The experiments demonstrated that a higher accuracy, lesser complexity and overheads were obtained from the developed framework in comparison with other HAR techniques.

Sharmin Majumder, et.al (2020) studied the application domains of HAR [7]. A survey of the papers was presented that made the deployment of vision and inactive sensing in a fused model for recognizing the activities of individuals. The surveyed papers were classified on the basis of fusion approaches, attributes and the multimodality datasets were also taken in account. Challenges and also the possible future directions were also defined in order to exploit the fusion of these two sensing modalities in the realistic circumstances. Chen Chen, et.al (2016) intended HAR system that was executed in real-time. A depth camera and an inertial sensor had utilized in this system at the same time on the basis of earlier constructed sensor fusion technique [28]. Thereafter, a decision-level fusion was carried out. A multimodal HAR data set which was available publicly had deployed for quantifying this system in real-time. For this purpose, a set of human actions was taken. The overall classification rate obtained from intended system was evaluated above 97%. The outcomes achieved in offline and real-time experiment represented that intended system was efficient.

Hang Yan, et.al (2020) set forth a compelling methodology with recognize the ceaseless Human recovery activity progressively [9]. In order to achieve this, the OpenPose and FCN were implemented. First of all, the Kalman filter was employed in the introduced approach for fusing the OpenPose so that the human targets were tracked and the two-dimensional poses action sequences were produced from the Red, Green, Blue (RGB) videos stream. After that, the segmented action sequence was extracted through sliding the window and the rectangular coordinates were transformed to relative coordinates from every frame of the human skeleton. The spatial-temporal attributes were extracted and the actions were identified by planning a one-dimensional fully convolutional network. The outcomes of experiment exhibited that the introduced approach was capable of recognizing the continuous rehabilitation actions online and an accuracy rate was computed 85.6%.

Naresh Kumar, et.al (2018) accentuated HAR dependent on skeletal and introduced a movement direction calculation strategy for which Fourier temporal representatives taken from the insertion of skeletal joints of human body were employed [10]. The human movement had considered as a direction of skeleton joints [10]. The human motion had taken in account as a path of skeletal joints for achieving this. The MSRAction3D benchmark   dataset was established for testing the presented method. This dataset categorized 3 action sets named AS1, AS2 and AS3 for carrying out the experiments. It was observed in the training and testing samples that the presented method provided the accuracy of 95.32 % while identifying the human action. The experiments authenticated that this method performed better than various other techniques.

Suraj Prakash Sahoo, et.al (2020) designed a DBiLSTM for sequential learning while establishing the temporal association among the action frames [11]. A pre-trained CNN was implemented for extracting the action information in every frame. The DHI was constructed by estimating and projecting the extracted information on the X-Y plane. The shape information was employed for the training of implemented model in the shape learning. The actions were identified from the query DHI images applying the fine-tuned network. The Data augmentation was adopted for avoiding the issue of overfitting of the network with the enlargement of training set virtually. The outcomes revealed that the designed work had provided superior performance in comparison with other existing algorithms with regard to accuracy and kappa parameter.

Carlos Roig, et.al (2019) presented a multi-modal approach that unified task centric features to recognize the activities of individuals [12]. This work had two aspects. The first was to establish a feature fusion block in which a gating system was deployed for carrying out attention over attributes from other domains. The second was to implement a pyramidal feature combination approach in which pairs of features taken from different tasks were integrated in hierarchical order with the help of preceding fusion block. The pyramid created the fine attributes which had applied in identifying the human action. A subgroup of the moments in time dataset was executed for authenticating the suggested technique. The results represented that this approach provided 35.43% precision.

Xueping Liu, et.al (2019) leveraged the Kinect abstraction technology and individual action tracing technology [13]. The Kinect SDK provided the deep data streams that assisted in extracting the detailed photographs ofa person body and info of skeletal data. In addition, the application of HAR was compared for comprehending the HAR. The intelligent surveillance platform was developed to validate the presented technique. The outcomes demonstrated that the presented technique performed well and provided fine application value.

eng-Lin Chiang, et.al (2018) set up a HAR framework dependent on vision to assist interaction between people and machines [14]. Initially, depth sensor and RGB camera were employed to capture depth images and color images at the same time with the utilization of Kinect 2.0. Then, the information of captured images was deployed in a color motion map and 3 depth motion maps. Their incorporation was done into one image for computing the corresponding HOG attributes. At last, these attributes of Histogram of Oriented Gradient were identified via SVM. The established system was had potential for recognizing 8 types of human actions. This system was tested using 3 data sets. The outcomes of experiment validated that the established system had offered efficiency and robustness.

Tin ZarWint Cho, et.al (2018) intended a system that emphasized on improving HAR with the implementation of skeletal features taken from Kinect sensor so that the discriminative features were achieved [15]. The static K-means algorithm was executed for clustering these features rather than applying the conventional K-means. The randomized initial centroids were mitigated for enhancing the accuracy of postures selection. The ANN using which the intended system was made wiser had carried out in order to label every posture. The HMM was implemented to recognize the human action on the basis of sequence of known poses for increasing the performance and accuracy. The intended system was capable of identifying the fundamental actions and UTKinect-Action3D, publicly available dataset was employed to compute this

system. The outcomes of experiment indicated that the static K-means provided higher accuracy as compared to the non-static K-means.

Peng Wang, et.al (2019) put forth a ConvNet dependent HAR algorithmic approach. The objective of this approach was to identify for the movement of human semaphore [16]. At first, data was gathered in 3 scenarios and the data enhancement was carried out using DCGAN for creating a dataset. Afterward, the complete convolution network was constructed by the means of 1*1 and 3*3 convolution kernels. The group convolution was executed for further compressing the model so that a novel model named HARNET was generated. It was evaluated in the experiments that the mAP of HARNET provided 94.36% performance on the DataSR dataset.

Sui Longfei, et.al (2018) proposed a strategy for perceiving the activities of human body using 3D information of human skeletal. For this purpose, Kinect depth image capturing device was used [17]. The places of 20 bone-focuses in the human body were caught; the typical points which were reasonable for the strategy had picked and the equivalent angle was computed through the Kinect in the suggested technique. This technique was adaptable for recognizing the human actions. The outcomes of experiment represented the feasibility and efficiency of suggested technique.

III. Summary

To perceive the human movement is viewed as exceptionally fundamental in people-to-people communication and social relations because of its inclination of giving data in regards to the character of a group, their personality and mental condition. The extraction of this data is exceptionally difficult. The major subject of study of the technical fields of computer vision and ML is the ability of an individual to trace the actions of others. The individuals' gestures are complex and also very dynamic.

References

1. Xinyi Liang, Hong-Bo Zhang, Yi-Xiang Zhang, Jin-Long Huang, "JTCR: Joint Trajectory Character Recognition for human action recognition", 2019, IEEE Eurasia Conference on IOT, Communication and Engineering (ECICE).

2. Qinkun Xiao, Yang Si, "Human action recognition using autoencoder", 2017, 3rd IEEE International Conference on Computer and Communications (ICCC).

3. Thien Huynh-The, Cam-Hao Hua, Nguyen Anh Tu, Jae-Woo Kim, Seung-Hwan Kim, Dong-Seong Kim, "3D Action Recognition Exploiting Hierarchical Deep Feature Fusion Model", 2020, 14th International Conference on Ubiquitous Information Management and Communication (IMCOM).

4. B Jagadeesh, Chandrashekar M Patil, Video based action detection and recognition human using optical flow and SVM classifier", 2016, IEEE International Conference on Recent Trends in Electronics, Information & Communication Technology (RTEICT).

5. Amir Nadeem, Ahmad Jalal, Kibum Kim, "Human Actions Tracking and Recognition Based on Body Parts Detection via Artificial Neural Network", 2020, 3rd International Conference on Advancements in Computational Sciences (ICACS)

6. Yun Han, Sheng-Luen Chung, Arul MuruganAmbikapathi, Jui-Shan Chan, Wei-You Lin, Shun-Feng Su, "Robust Human Action Recognition Using Global Spatial-Temporal Attention for Human Skeleton Data", 2018, International Joint Conference on Neural Networks (IJCNN)

7. Sharmin Majumder, Nasser Kehtarnavaz, "Vision and Inertial Sensing Fusion for Human Action Recognition: A Review", 2020, IEEE Sensors Journal.

8. Chen Chen, Roozbeh Jafari, Nasser Kehtarnavaz, "A Real-Time Human Action Recognition System Using Depth and Inertial Sensor Fusion", 2016, IEEE Sensors Journal.

9. Hang Yan, Beichen Hu, Gang Chen, E. Zhengyuan, "Real-Time Continuous Human Rehabilitation Action Recognition using OpenPose and FCN", 2020, 3rd International Conference on Advanced Electronic Materials, Computers and Software Engineering (AEMCSE).

10. Naresh Kumar, Nagarajan Sukavanam, "Human Action Recognition from Motion Trajectory using Fourier Temporal Features of Skeleton Joints", 2018, International Conference on Advances in Computing and Communication Engineering (ICACCE).

11. Suraj Prakash Sahoo, Samit Ari, Kamalakanta Mahapatra, Saraju P. Mohanty, "HAR-Depth: A Novel Framework for Human Action Recognition Using Sequential Learning and Depth Estimated History Images", 2020, IEEE Transactions on Emerging Topics in Computational Intelligence.

12 Carlos Roig, Manuel Sarmiento, David Varas, Issey Masuda, Juan Carlos Riveiro, ElisendaBou-Balust, "Multi-Modal Pyramid Feature Combination for Human Action Recognition", 2019, IEEE/CVF International Conference on Computer Vision Workshop (ICCVW).

13. Xueping Liu, Yibo Li, Youru Li, Shi Yu, Can Tian, "The Study on Human Action Recognition with Depth Video for Intelligent Monitoring", 2019, Chinese Control And Decision Conference (CCDC).

14. Meng-Lin Chiang, Jian-Kai Feng, Wen-Lin Zeng, Chiung-Yao Fang, Sei-Wang Chen, "A Vision-Based Human Action Recognition System for Companion Robots and HumanInteraction", 2018, IEEE 4th International Conference on Computer and Communications (ICCC).

15. Tin ZarWint Cho, May Thu Win, Aung Win, "Human Action Recognition System based on Skeleton Data", 2018, IEEE International Conference on Agents (ICA).

16. Peng Wang, Yuliang Yang, Wanchong Li, Linhao Zhang, Mengyuan Wang, Xiaobo Zhang, Mengyu Zhu, "Research on Human Action Recognition Based on Convolutional Neural Network", 2019, 28th Wireless and Optical Communications Conference (WOCC).

17. Sui Longfei, Miao Lanfang, Li Zhifei, "Human Body Action Recognition Based on Bone Information by Kinect", 2018, International Computers, Signals and Systems Conference (ICOMSSC).

18. A. K. Phulre, M. Kamble and S. Phulre, "Content Management Systems hacking probabilities for Admin Access with Google Dorking and database code injection for web content security," 2nd International Conference on Data, Engineering and Applications (IDEA), 2020,pp.1-5,doi: 10.1109/IDEA49133.2020. 9170655.

19. S. Phulre, P. Gautam and S. K. Mishra, "Implementation of trusted multitier method for intrusion detection in mobile ad hoc networks with DSR algorithm," 2014 Science and Information Conference, 2014, pp. 666-673, doi: 10.1109/SAI.2014.6918258.

20. P. S. Tomar, S. Phulre and M. Shrivastava, "Impact of Varying Physical Links on QoS of Virtual Topology Management in WDM Networks," 2021 10th IEEE International Conference on Communication Systems and Network Technologies (CSNT), 2021, pp. 74-80, doi: 10.1109/CSNT51715.2021.9509699.

21. S. Sangulle, Sunil Phulre, "Sentiment Detection Using Fish Optimization Genetic Algorithm", nternational Journal of Engineering Trends and Technology Volume 68 Issue 12, 140-145, December 2020 ISSN: 2231 – 5381 /doi:10.14445/22315381/IJETT-V68I12P223.

22. Sunil Phulre, "WSN ENERGY OPTIMIZATION USING INTELLIGENT WATER DROP GENETIC ALGORITHMS", International Journal of Advanced Research in Engineering a nd Technology (IJARET) Volume 11, Issue 12, December 2020, pp.2449-2462, DOI: 10.34218/IJARET.11.12.2020.231, ISSN Print: 0976-6480 and ISSN Online: 0976-6499.

23. Sukhlal Sangule,, Sunil Phulre "A Survey on Digital Content Sentiment Features and Techniques", International Journal of Computer Sciences and Engineering, Vol.8, Issue.3, Mar 2020 E-ISSN: 2347-2693.