International Journal of Scientific Research in Computer Science, Engineering and Information Technology
ISSN : 2456-3307 (www.ijsrcseit.com)
doi : https://doi.org/10.32628/CSEIT2390379

Human Activity Recognition Using Deep Learning : A Survey

Bhushan Marutirao Nanche[1], Dr. Hiren Jayantilal Dand[2], Dr. Bhagyashree Tingare[3]
[1]Research Scholar, JJTU, Jhunjhunu, Rajasthan, India
[2]Research Guide, JJTU Jhunjhunu, Rajasthan, India
[3]Research Co-Guide, DYPCOE, Pune, Maharashtra, India

ARTICLEINFO

Article History:

Accepted: 10 June 2023
Published: 23 June 2023

Publication Issue
Volume 9, Issue 3
May-June-2023

Page Number
605-610

ABSTRACT

With the use of deep learning algorithms from artificial intelligence (AI), several types of research have been conducted on video data. Object localization, behaviour analysis, scene understanding, scene labelling, human activity recognition (HAR), and event recognition make up the majority of them. Among all of them, HAR is one of the most difficult jobs and key areas of research in video data processing. HAR can be used in a variety of fields, including robotics, human-computer interaction, video surveillance, and human behaviour categorization. This research seeks to compare deep learning approaches on several benchmark video datasets for vision-based human activity detection. We suggest a brand-new taxonomy for dividing up the literature into CNN- and RNN-based methods. We further categorise these approaches into four subgroups and show several methodologies, their effectiveness, and experimental datasets. To illustrate the development of HAR techniques, a brief comparison is also provided with the handcrafted feature-based approach and its merger with deep learning. Finally, we go over potential future research areas and some unresolved issues with recognising human activities. This survey's goal is to present the most recent developments in HAR techniques for vision-based deep learning using the most recent literature analysis.

Keywords : Human Activity Recognition (HAR), Deep Learning, Artificial Intelligence, Computer Vision

I. INTRODUCTION

In the age of the smart city, video monitoring has become essential for improving the quality of life and creating secure zones. Surveillance cameras are often mounted at a specific distance for optimum area coverage. As a result, better analysis and a deeper comprehension of films are absolutely necessary, which has a significant impact on the security system. A video data-driven system is beneficial to the healthcare, transportation, manufacturing, educational, and retail sectors. Every camera feed's goal is to identify the specific incident, such as identifying suspicious activity [1] at an airport, bus stop, or train station, unusual activity [2] at a public gathering, or an unusual pattern of behaviour by factory workers [3].

These are the select few illustrative areas when acknowledging human action is particularly desirable. In HAR-based systems, abnormal activity typically triggers an alert to the control room. Instead of sitting in front of the camera feed and observing what is happening every second, it is imperative to be aware of certain specific items in such situations.

To accurately represent human actions and their interactions from an unheard-of data sequence is the main goal of human activity recognition. Due to a number of issues such as shifting backgrounds and poor video quality, it is frequently difficult to reliably identify human activity from video data. The two primary issues that are raised by different human activity identification systems are: "Which action is performed?" (also known as the action recognition task) and "Where exactly in the video?" (also known as the localization problem). The collections of photographs are known as frames. An action recognition task's main goal is to analyse the input video clips in order to identify the subsequent human activities.

Human behaviour imitates their patterns, therefore each human action is distinct, making it difficult to identify. Another difficult problem is creating such a deep learning-based model to forecast human behaviour within acceptable benchmark datasets for assessment. The enormous success of the ImageNet [4] dataset for image processing has led to the publication of multiple benchmark action recognition datasets [5] [6] to further this field's study. Similar to image processing, let's consider how much computing power and input parameters are needed to train a deep learning model for video data processing.

II. LITERATURE SURVEY

Techniques for recognising human activity range from manually constructed feature-based methods to cutting-edge AI-based deep learning methods. Human activity recognition has been surveyed by authors [7] who divided the study's scope into data modalities and their applications. The study is further subdivided based on model development techniques and different HAR activities. The authors look at the unimodal and multimodal HAR approaches in the major classification. Space-time, stochastic, rule-based, and shape-based models are grouped under Unimodal categories. The emotive, behavioural, and social networking subcategories of human activity are simultaneously listed by multimodal.

HAR for production and logistics was the subject of a thorough literature evaluation conducted by Reining et al. [8]. The state-of-the-art HAR methods, statistical pattern recognition, and deep architectures are all covered in-depth in this examination. The industrial applications of this work are advantageous. Vision-based human action recognition was surveyed by Beddiar et al. [9], who divided the entire study into the following categories: A handcrafted feature and feature learning-based method were used, and the authors described the different techniques and the specifics of how they should be put into practise. The authors also draw attention to relevant material that supports HAR methods at the minute level and is based on categories of human activity, including elementary human activities, gestures, behaviours, interactions, group actions, and events.

Similarly to this, Zhu et al. [10] looked at both custom-made and learning-based methods for action recognition. Unlike [11], the authors briefly describe the rise of HAR's deep learning approaches up until 2016 after evaluating the handcrafted method's limitations. The development of cutting-edge activity recognition methodologies in terms of activity representation and HAR classification algorithms is the focus of a review by Zhang et al. [12]. The classification of classification approaches is based on template, discriminative, and generative models, while the classification of representation elements is based on global, local, and advanced depth-based. The HAR dataset and the succinctly described models demonstrate performance accuracy in the experimental results. The study uses only HAR categorization techniques that are current as of 2017.

Another study conducted in the same year by Herath et al. [13] reveals a comparable investigation that was started with the inventor of the HAR methodology, a handcrafted feature-based approach to deep learning-based methodologies. The previous surveys lacked a thorough presentation of deep learning techniques that mapped with HAR datasets, but this study does just that. However, it covers the body of literature up to 2016, thus researchers must have access to developments made after then. These authors' well-defined predictions for the future are a great incentive to adopt them within the scholarly community.

In their study [14], Koohzadi et al. explore the use of deep learning in HAR image and video processing. Supervised-deep generative, Supervised-deep discriminative, Unsupervised-deep, Semi-supervised-deep, and Hybrid models are the five categories under which the whole technique is divided. The advantages and advice for selecting a deep learning model for HAR in the aforementioned five areas is one distinctive point made in this survey.

The author also covered deep learning methods for spatiotemporal representation, which involves expanding typical 2D image processing to include time as a third dimension. In their article from 2018, Nweke et al. [15] provide a thorough analysis of deep learning techniques for mobile and wearable sensor-based HAR. Methods are categorised in generative, discriminative, and hybrid ways by outlining their benefits and drawbacks. Instead of using activity recognition datasets based on vision, this study assesses deep learning techniques using mobile sensor-based human activity recognition datasets. The authors contrast standard feature learning with deep learning feature representation techniques. The difficulties of using sensor networks for HAR are also covered.

The survey by Zhang et al. [16] demonstrates advancements in human-object interaction identification techniques, action feature representation techniques based on deep learning, and action features for depth and RGB data. This study differs from earlier work in that its primary contribution thoroughly explains the handcrafted action feature for RGB, depth, and skeleton data. The HAR datasets, which were available until 2018, are also useful for discussing the performance evaluations of deep learning systems.

Singh et al. [17] ran a survey to help researchers find the best HAR dataset for benchmarking their algorithms. The current HAR dataset divides images into RGB and RGBD (depth) categories. In terms of lighting variation, annotation, occlusions, perspective variation, and fusion modalities, challenges with these datasets are also highlighted. The RGB-Depth sensor-based HAR survey presented by Liu et al. [18] discusses handcrafted and learning-based characteristics. Within the three subcategories of Depth-based methods, Skeleton-based methods, and Hybrid feature-based methods, this survey presents a fresh taxonomy for both methods. This survey briefly evaluates the accuracy results of the deep learning approach on RGB-D-based human activity datasets. For RGB-Depth sensor-based HAR, difficulties and future research are also mentioned.

The authors of the survey by Hussain et al. [19] examine many HAR topics with a main emphasis on device-free approaches, particularly RFID. Based on the relevant research, the author suggests a new taxonomy with three sub-areas: action-based, motion-based, and interaction-based. The most recent HAR techniques are described under each of these sub-themes, which are further separated into ten separate subjects.

The authors of a related survey by Dang et al. [20] discussed both the sensor-based and vision-based HAR approaches in detail. Each group is further divided into smaller units that carry out particular tasks, such as data gathering, pre-processing techniques, feature engineering, and training. Along with difficulties and potential directions, deep learning HAR approaches are also briefly described.

The kinetics-based literature, which discusses the use of the Kinect camera for data collecting and deep learning algorithms for activity recognition, is used by Wang et al. [21]. Using six kinetics-based datasets, the

authors reviewed 10 Kinect-based algorithms for cross-subject action detection and cross-view action recognition. For researchers using the Microsoft Azure Kinect Developer Kit to create HAR models for real-time applications, this survey represents a fresh source. The difficulties with HAR techniques and datasets were resolved by the authors Jegham et al. [22]. They concentrated on conducting surveys to look into an overview of the current approaches in light of the many kinds of problems illustrated in the literature. This survey encourages researchers in computer vision to identify the major difficulties in HAR to decide on future research to address these practical applications. A study by Majumder et al. [23] provided the literature evidence of the fusion of vision and inertial sensing. This information was used to increase the accuracy of the HAR system.

It is decided to conduct the first survey of its kind in this area using fusion techniques, features, classifiers, and multimodality datasets. Network-based, motion-based, multiple instances are learning-based, dictionary-based, and histogram-based approaches are the categories that Ozyer et al. [24] use to classify the existing HAR methods. Additionally, they contrasted the outcomes on HAR datasets.

The author [25] has reviewed a number of supervised and unsupervised machine learning methods for identifying human behaviour. The support vector machine (SVM), the hidden markov model (HMM), and the neural network are examples of supervised learning approaches (classification and regression) where the authors documented the influential literature for abnormal behaviour and activity detection. Whereas, the author reported object trajectory analysis and pixel-based features for aberrant behaviour detection in video sequences under the category of unsupervised learning approach (Clustering) for anomaly detection. For the purposes of track analysis, moving hands, multiple objects, behaviour analysis, walking, running, and cycling on the highway, there are many different types of clustering methods, including partition-based clustering, hierarchical, density-based latent, and Gaussian technique.

III. CONCLUSION

In conclusion, we found that the majority of the surveys include a taxonomy for categorising HAR approaches for reasons of comparison. We also observed a wide range of HAR approaches in the comparative surveys, including dataset-based, input-type-based, HAR real-world challenge-based, and learning-based approaches. In this context, we declare that the strategy we used for this survey is a learning-based strategy, and we suggest a brand-new taxonomy of research based on the design of current deep learning algorithms. More than 25 modern deep learning-based algorithms have been covered, and their performance on benchmark HAR datasets for vision-based applications has been provided.

IV. REFERENCES

[1]. Chen, D., P. Wang, L. Yue, Y. Zhang, and T. Jia. 2020. Anomaly Detection in Surveillance Video Based on Bidirectional Prediction. Image and Vision Computing 98:103915. doi:10.1016/j.imavis.2020.103915.

[2]. Wang, S., Y. Liu, J. Wang, S. Gao, and W. Yang. 2021. A Moving Track Data-Based Method for Gathering Behavior Prediction at Early Stage. Applied Intelligence 51 (11):8498–518. doi: 10.1007/s10489-021-02244-2.

[3]. Wenjin, T., Z. Hao Lai, M. C. Leu, and Z. Yin. 2018. Worker Activity Recognition in Smart Manufacturing Using IMU and SEMG Signals with Convolutional Neural Networks. Procedia Manufacturing 26:1159–66. Elsevier B.V. doi:10.1016/j.promfg.2018.07.152.

[4]. Deng, J., R. S. Wei Dong, L. Li-Jia, L. Kai, and L. Fei-Fei. 2009. ImageNet: A Large-Scale Hierarchical Image Database. 2009 IEEE Conference on Computer Vision and Pattern

Recognition, 248–55. doi: 10.1109/cvprw.2009.5206848.

[5]. Kay, W., J. Carreira, K. Simonyan, B. Zhang, C. Hillier, S. Vijayanarasimhan, F. Viola, T. Green, T. Back, P. Natsev, M. Suleyman, and A. Zisserman. 2017. The Kinetics Human Action Video Dataset. ArXiv.

[6]. Soomro, K., A. Roshan Zamir, and M. Shah. 2012. "UCF101: A Dataset of 101 Human Actions Classes from Videos in the Wild". November. http://arxiv.org/abs/1212.0402.

[7]. Michalis, V., C. Nikou, and I. A. Kakadiaris. 2015. A Review of Human Activity Recognition Methods. Frontiers Robotics AI 2 (NOV):1–28. doi:10.3389/frobt.2015.00028.

[8]. Christopher, R., F. Niemann, F. Moya Rueda, G. A. Fink, and M. ten Hompel. 2019. Human Activity Recognition for Production and Logistics-a Systematic Literature Review. Information (Switzerland) 10 (8):1–28. doi:10.3390/info10080245.

[9]. Beddiar, D. R., B. Nini, M. Sabokrou, and A. Hadid. 2020. Vision-Based Human Activity Recognition: A Survey. Multimedia Tools and Applications 79 (41–42):30509–55. doi:10.1007/s11042-020-09004-3.

[10]. Zhu, F., L. Shao, J. Xie, and Y. Fang. 2016. From Handcrafted to Learned Representations for Human Action Recognition: A Survey. Image and Vision Computing 55:42–52. doi:10.1016/ j.imavis.2016.06.007.

[11]. Beddiar, D. R., B. Nini, M. Sabokrou, and A. Hadid. 2020. Vision-Based Human Activity Recognition: A Survey. Multimedia Tools and Applications 79 (41–42):30509–55. doi:10.1007/s11042-020-09004-3.

[12]. Zhang, S., Z. Wei, J. Nie, L. Huang, S. Wang, and Z. Li. 2017. A Review on Human Activity Recognition Using Vision-Based Method. Journal of Healthcare Engineering 2017:1–31. doi:10.1155/2017/3090343.

[13]. Herath, S., M. Harandi, and F. Porikli. 2017. Going Deeper into Action Recognition: A Survey. Image and Vision Computing 60:4–21. doi:10.1016/j.imavis.2017.01.010.

[14]. Koohzadi, M., and N. Moghadam Charkari. 2017. Survey on Deep Learning Methods in Human Action Recognition. IET Computer Vision 11 (8):623–32. doi:10.1049/iet-cvi. 2016.0355.

[15]. Nweke, H. F., Y. Wah Teh, M. Ali Al-garadi, and U. Rita Alo. 2018. Deep Learning Algorithms for Human Activity Recognition Using Mobile and Wearable Sensor Networks: State of the Art and Research Challenges. Expert Systems with Applications 105:233–61. doi:10.1016/j. eswa.2018.03.056.

[16]. Zhang, H.-B., Y.-X. Zhang, B. Zhong, Q. Lei, L. Yang, D. Ji-Xiang, and D.-S. Chen. 2019. A Comprehensive Survey of Vision-Based Human Action Recognition Methods. Mpdi. doi:10.3390/s19051005.

[17]. Singh, T., and D. Kumar Vishwakarma. 2019. Video Benchmarks of Human Action Datasets: A Review. Artificial Intelligence Review 52 (2):1107–54. doi: 10.1007/s10462-018-9651-1.

[18]. Liu, B., H. Cai, J. Zhaojie, and H. Liu. 2019. RGB-D Sensing Based Human Action and Interaction Analysis: A Survey. Pattern Recognition 94:1–12. doi: 10.1016/j. patcog.2019.05.020.

[19]. Zawar, H., Q. Z. Sheng, and W. Emma Zhang. 2020. A Review and Categorization of Techniques on Device-Free Human Activity Recognition. Journal of Network and Computer Applications 167:102738. December 2019. doi:10.1016/j.jnca.2020.102738.

[20]. Minh Dang, L., K. Min, H. Wang, M. Jalil Piran, C. Hee Lee, and H. Moon. 2020. Sensor-Based and Vision-Based Human Activity Recognition: A Comprehensive Survey. Pattern Recognition 108:107561. doi:10.1016/j.patcog.2020.107561.

[21]. Lei, W., D. Q. Huynh, and P. Koniusz. 2020. A Comparative Review of Recent Kinect-Based Action Recognition Algorithms. IEEE

Transactions on Image Processing 29:15–28. doi:10.1109/TIP.2019.2925285.

[22]. Jegham, I., A. Ben Khalifa, I. Alouani, and M. Ali Mahjoub. 2020. Vision-Based Human Action Recognition: An Overview and Real World Challenges. Forensic Science International: Digital Investigation 32:200901. doi:10.1016/j.fsidi.2019.200901.

[23]. Majumder, S., and N. Kehtarnavaz. 2021. Vision and Inertial Sensing Fusion for Human Action Recognition: A Review. IEEE Sensors Journal 21 (3):2454–67. doi:10.1109/ JSEN.2020.3022326.

[24]. Özyer, T., A. Duygu Selin, and R. Alhajj. 2021. Human Action Recognition Approaches with Video Datasets—A Survey. Knowledge-Based Systems 222:106995. doi:10.1016/j. knosys.2021.106995.

[25]. Verma, K. K., B. Mohan Singh, and A. Dixit. 2022. A Review of Supervised and Unsupervised Machine Learning Techniques for Suspicious Behavior Recognition in Intelligent Surveillance System. International Journal of Information Technology (Singapore) 14 (1):397–410. doi:10.1007/s41870-019-00364-0.

Cite this article as :

Bhushan Marutirao Nanche, Dr. Hiren Jayantilal Dand, Dr. Bhagyashree Tingare, "Human Activity Recognition Using Deep Learning : A Survey", International Journal of Scientific Research in Computer Science, Engineering and Information Technology (IJSRCSEIT), ISSN : 2456-3307, Volume 9, Issue 3, pp.605-610, May-June-2023. Available at doi : https://doi.org/10.32628/CSEIT2390379
Journal URL : https://ijsrcseit.com/CSEIT2390379