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

Harnessing Convolutional Neural Networks and Transfer Learning to Perform Vision-Oriented Activity Recognition of Humans

Dr. Sunil Bhutada[1], B. Yeshwanth Raj[2]

[1]HOD & Professor, Department of IT, Sreenidhi Institute of Science & Technology, Yamnampet, Ghatkesar, Hyderabad, India

[2]MTech Student, Department of IT, Sreenidhi Institute of Science & Technology, Yamnampet, Ghatkesar, Hyderabad, India

Article Info

Publication Issue :
Volume 9, Issue 1
January-February-2023

Page Number : 04-08

Article History
Accepted: 01 Jan 2023
Published: 10 Jan 2023

ABSTRACT

Since the debut of the Internet of Things, there has always been a lot of noteworthy evolution in the aspect of "Human Activity recognition". Recognition of activities of humans possesses its own connotation and purpose and can be employed in diverse range of disciplines featuring medical assistance, nefarious activities, and espionage. It's possible that it could be critically pertinent in order to undergo ample amount of criminological investigations. To anticipate various human behaviors, a myriad of machine learning techniques are used. However, deep learning models have trounced standard machine learning strategies. Convolutional Neural Networks (CNN), a type of deep learning model, could very well heuristically extract the features and drastically cut overall processing expenditure. The action recognition kinetics dataset can be used to predict human activities using the CNN model. Here, we use transfer learning specifically for visual categorization problems.

Keywords: Human Activity Recognition, Criminological Investigations, Convolutional Neural Networks, Transfer Learning.

I. INTRODUCTION

Human activity recognition (HAR) is very much an interesting research topic which scrutinizes diverse range of activities from video sequences or still images, as well as in an assortment of other ways in which AI, ML, and IOT can be employed. It has gained prominence and relevance as a corollary of its properties and potential applications in domains namely in elder care and health assistance, video monitoring, and many others. It is predominantly deployed in a wide variety of disciplines where vital information about one's cognitive functioning and style of living is considered necessary. There are umpteen investigations and research studies based on HAR that are currently underway and that are also have previously been performed. Some of the prevailing systems employed in the field of HAR are wearable or non-wearable based; the former requires wearable tech with sensor systems that are or can be hooked up to human bodies like fitness bands, activity trackers and many others, whereas the latter literally

does not necessitate any sensors that must be attached to humans or devices to be carried for activity recognition. Sensory and vision-focused HAR systems fall within the category of non-wearable systems in this context.

Vision-focused techniques leverage video content and picture frames from depth or thermo graphic cameras to forecast a human activity, sensor-based techniques can use passive infrared sensors and radio frequency signals from radio frequency identification sensors to identify a human activity. Non-wearable systems like sensor-based systems can or cannot achieve great accuracy and wearable-oriented systems often necessitate the usage of expensive hardware. As a result of the shortcomings of the aforementioned systems, vision-oriented systems, which may be easily deployed through the use of videos, have become increasingly important in recent years.

Although difficult, identifying human activity in a streaming video can be done with ease. We mainly make use of RGB or a depth image in vision-oriented methods. No equipment is needed to be carried by the user, and no sensors are required to be placed over a human body. As a result, this methodology is currently being applied in a humongous way.

In recent years, deep learning techniques have supplanted almost all of the machine learning algorithms utilized in vision-oriented applications. Here, convolutional neural networks, which are one of the diverse accessions of deep learning approaches, were applied by us. They are widely and frequently utilized within fields pertaining to computer vision. They are able to generate features intuitively and greatly simplify the work of extracting features, which often demands for subject-specific knowledge. Here, transfer learning is employed in order to perform tasks of visual categorization.

II. EXISTING SYSTEM

The current technology is built on a system for recognizing human activities by means of wearable devices. The usage of various sensors by wearable systems necessitates constant contact with the human body in order to monitor the actions being performed by the user. A smart watch/band is one amongst the ample amount of examples of a wearable-based HAR system because of its ability to perform fitness and activity tracking as well as sleep monitoring.

2.1. DISADVANTAGES OF EXISTING SYSTEM

- Sensors and other devices take some time in order to get deployed .
- Wearable sensors and gadgets can cost a lot of money.

III. PROPOSED SYSTEM

Here, we deploy a vision-focused HAR system to classify human actions using videos and image frames from depth or thermo graphic cameras. The RGB or depth image is used in the vision-oriented strategy. As a result of this method's improved exposure today, it is used in a variety of applications. To facilitate the operation of the foregoing method, we additionally applied Convolutional Neural Networks, a common deep learning methodology. According to CNN, their applications are relevant to fields relating to computer vision. Using CNN's in this case has a number of benefits, one of which is their capacity to produce features automatically. This skill makes the work of extracting characteristics, which typically requires subject-oriented proficiency, much simpler. The proposed method can be used in the main fields such as criminological investigations and many more etc.

3.1 ADVANTAGES OF PROPOSED SYSTEM

- Uses pictures and videos rather than sensors or other equipment.
- Reduces the use of expensive sensors to a certain extent.

Algorithm: Convolutional Neural Networks (CNN), VGG-16 (also called Oxford Net)

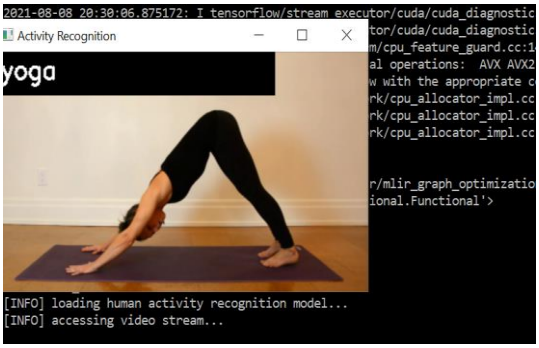
IV. EXPERIMENTAL SETUP

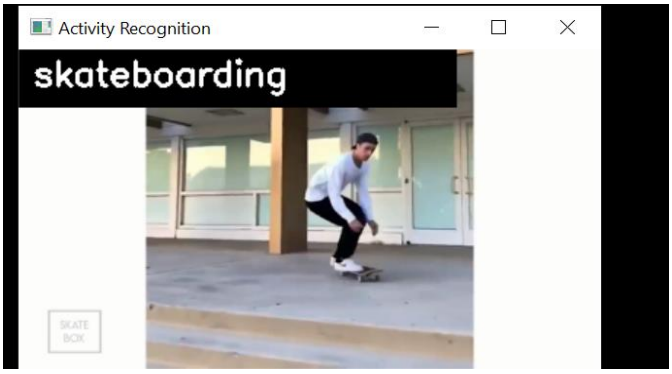
We made a use of convolutional neural networks for activity recognition of various people by means of action recognition kinetics dataset because cnn's make the process of extraction of features easier that mainly require certain subject oriented knowledge. We extracted frames from each activity shown in the video. On the other hand, we specifically used transfer learning to undergo visual categorization tasks. We applied CNN models such as VGG-16, Inspect 32, etc to classify human activities.

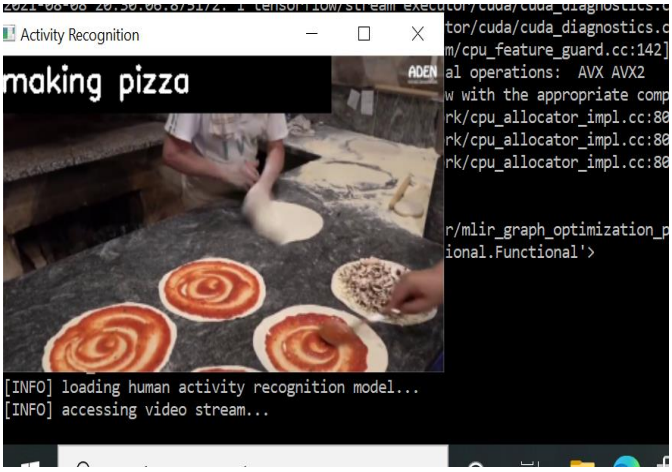
We used python 3.7 to install packages of open source libraries such as opencv, scikit learn and tensor flow which we require to implement CNN and transfer learning and also to run the code.

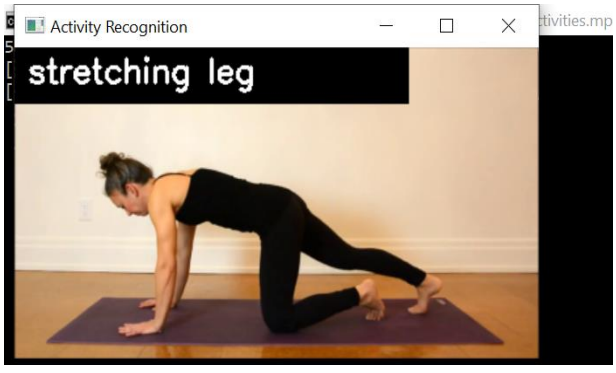
V. RESULTS

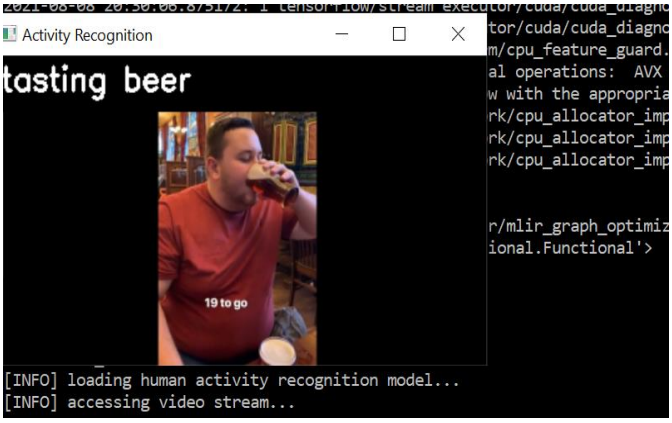
In the output of our proposed system we get to see the activities performed by various people in a video that will stream after we run the program. The activities performed by them are displayed in the form of written text on the top left corner of the video. This kind of method is useful where knowing the activities performed by people becomes somewhat difficult to understand and can serve as useful option in crime investigation.

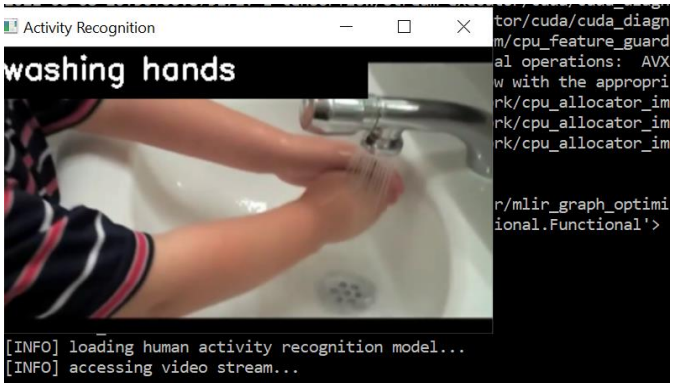












VI. CONCLUSION

We hereby conclude that our proposed method is helpful in the areas where video surveillance has its utmost importance. Vision based human activity recognition makes the process of activity recognition by means of streaming and recorded video that are generally captured by cctv's or cameras. It is useful where knowing the activities performed by people becomes somewhat difficult to understand and can serve as useful option in crime investigation etc. It reduces the need for installing any kind of wearable or non wearable kinds of sensors that are generally used for human activity recognition as we can easily perform activity recognition process by means of videos.

VII. REFERENCES

[1]. B. Bhandari, J. Lu, X. Zheng, S. Rajasegarar, and C. Karmakar, "Noninvasive sensor based automated smoking activity detection," in Proceedings of Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC). IEEE, 2017, pp. 845–848.

[2]. L. Yao, Q. Z. Sheng, X. Li, T. Gu, M. Tan, X. Wang, S. Wang, and W. Ruan, "Compressive representation for device-free activity recognition with passive rfid signal strength," IEEE Transactions on Mobile Computing, vol. 17, no. 2, pp. 293–306, 2018.

[3]. I. Lillo, J. C. Niebles, and A. Soto, "Sparse composition of body poses and atomic actions for human activity recognition in rgb-d videos," Image and Vision Computing, vol. 59, pp. 63–75, 2017.

[4]. W. Zhu, C. Lan, J. Xing, W. Zeng, Y. Li, L. Shen, and X. Xie, "Co-occurrence feature learning for skeleton based action recognition using regularized deep lstm networks," in Thirtieth AAAI Conference on Artificial Intelligence, 2016.

[5]. C. Szegedy, W. Liu, Y. Jia, P. Sermanet, S. Reed, D. Anguelov, D. Erhan, V. Vanhoucke, and A. Rabinovich, "Going deeper with convolutions," in Proceedings of the IEEE conference on computer vision and pattern recognition, 2015, pp. 1–9.

[6]. J. Deng, W. Dong, R. Socher, L. Li, Kai Li, and Li Fei-Fei, "ImageNet: A large-scale hierarchical image database," in Proceedings of IEEE Conference on Computer Vision and Pattern Recognition, June 2009, pp. 248–255.

[7]. A. Jalal, N. Sarif, J. T. Kim, and T.-S. Kim, "Human activity recognition via recognized body parts of human depth silhouettes for residents monitoring services at smart home," Indoor and built environment, vol. 22, no. 1, pp. 271–279, 2013.

[8]. K. Simonyan and A. Zisserman, "Two-stream convolutional networks for action recognition in videos," in Advances in neural information processing systems, 2014, pp. 568–576.

[9]. G. Gkioxari, R. Girshick, and J. Malik, "Contextual action recognition with r* cnn," in Proceedings of the IEEE international conference on computer vision, 2015, pp. 1080–1088.

[10]. L. Wang, Y. Xiong, Z. Wang, and Y. Qiao, "Towards good practices for very deep two-

stream convnets," arXiv preprint arXiv:1507.02159, 2015.

[11]. D. Tran, L. Bourdev, R. Fergus, L. Torresani, and M. Paluri, "Deep end2end voxel2voxel prediction," in Proceedings of the IEEE conference on computer vision and pattern recognition workshops, 2016, pp. 17–24.

[12]. P. Wang, W. Li, J. Wan, P. Ogunbona, and X. Liu, "Cooperative trainingof deep aggregation networks for rgb-d action recognition," in Thirty-Second AAAI Conference on Artificial Intelligence, 2018.

[13]. S. Ji, W. Xu, M. Yang, and K. Yu, "3d convolutional neural networks for human action recognition," IEEE transactions on pattern analysis and machine intelligence, vol. 35, no. 1, pp. 221–231, 2013.

[14]. P. Khaire, P. Kumar, and J. Imran, "Combining cnn streams of rgb-d and skeletal data for human activity recognition," Pattern Recognition Letters, vol. 115, pp. 107–116, 2018.

AUTHORS PROFILE

Dr. Sunil Bhutada[1] is a professor and Head of the department of Information Technology, SNIST, Hyderabad. He completed his B.E in Computer science from Amravati University, Pusad, M.Tech in Software Engineering from JNTU, Hyderabad and successfully completed his Ph.D. from JNTU Hyderabad in 2017 in the area of data mining, topic modeling. He is having teaching and research experience of more than 24 years. He published research papers which have reached a count of 98 among which 14 are Scopus Indexed journals and one in Taylor and Francis journal. 8 papers of his got published in international conferences and 4 papers got published in national conferences.

Mr. B. Yeshwanth Raj[2], is a student who is doing M.Tech course in Computer networks and Information security under the department of Information technology in Sreenidhi Institute of science and technology (SNIST). He has a B.Tech degree in Electronics and communication engineering from CMR college of engineering and technology. He completed his schooling from DAV Public school, Safilguda, Hyderabad.

Cite this article as :

Dr. Sunil Bhutada, B. Yeshwanth Raj, "Harnessing Convolutional Neural Networks and Transfer Learning to Perform Vision-Oriented Activity Recognition of Humans ", International Journal of Scientific Research in Computer Science, Engineering and Information Technology (IJSRCSEIT), ISSN : 2456-3307, Volume 9 Issue 1, pp. 04-08, January-February 2023.

Journal URL : https://ijsrcseit.com/CSEIT2390111