JOURNAL OF NORTHEASTERN UNIVERSITY

Volume 26 Issue 02, 2023 ISSN: 1005-3026 https://dbdxxb.cn/ Original Research Paper

LEVERAGING CNN AND TRANSFER LEARNING FOR VISION BASED HUMAN ACTIVITY RECOGNITION

[1].Dr.Abdul Raheem, [2]. S. Bhavana Sree, [3]. S. Bhavana, [4].Y. Sindhura

[1].Professor, [2,3&4].UG Scholar

Department of ECE, Malla Reddy Engineering College for Women, Hyderabad

ABSTRACT: With the advent of the Internet of Things (IoT), there have been significant advancements in the area of human activity recognition (HAR) in recent years. HAR is applicable to wider application such as elderly care, anomalous behaviour detection and surveillance system. Several machine learning algorithms have been employed to predict the activities performed by the human in an environment. However, traditional machine learning approaches have been outperformed by feature engineering methods which can select an optimal set of features. On the contrary, it is known that deep learning models such as Convolutional Neural Networks (CNN) can extract features and reduce the computational cost automatically. In this paper, we use CNN model to predict human activities from Wiezmann Dataset. Specifically, we employ transfer learning to get deep image features and trained machine learning classifiers. Our experimental results showed the accuracy of 96.95% using VGG16. Our experimental results also confirmed the high performance of VGG-16 as compared to rest of the applied CNN models.

Index Terms—Human activity recognition, sensing technology, depth sensor, wearable devices, RGB camera, Kinect

I. INTRODUCTION Human Activity Recognition (HAR) is one of the active research areas in computer vision as well as human computer interaction [1]–[3]. However, it remains a very complex task, due to unresolvable challenges such as sensor motion, sensor placement, cluttered background, and inherent variability in the way activities are conducted by different human [4], [5]. In this paper, a total of thirty-two recent research papers on sensing technologies used in HAR are reviewed. The most commonly employed sensing technologies in HAR system regardless of the computational models or classification algorithms are analyzed. The pros and cons of each sensing technology have been discussed. This paper is concluded with some challenges for the most sophisticated sensing technologies. Human activity recognition (HAR) is an active research area because of its applications in elderly care, automated homes and surveillance system. Several studies has been done on human activity recognition in the past. Some of the existing work are either wearable based [1] or non-wearable based [2] [3]. Wearable based HAR system make use of wearable sensors that are attached on the human body. Wearable based HAR system are intrusive in nature. Non-wearable based HAR system do not require any sensors to attach on the human or to carry any device for activity recognition. Non-wearable based approach can be further categorised into sensor based [2] and vision-based HAR systems [3]. Sensor based technology use RF signals from sensors, such as RFID, PIR sensors and WiFi signals to detect human activities. Vision based technology use videos, image frames from depth cameras or IR cameras to classify human activities. Sensor based HAR

Submitted: 22/04/2023 **Accepted**: 10/05/2023

system are non-intrusive in nature but may not provide high accuracy. Therefore, vision-based human activity recognition system has gained significant interest in the present time. Recognising human activities from the streaming video is challenging. Video-based human activity recognition can be categorised as marker-based and vision-based according to motion features [4]. Marker-based method make use of optic wearable markerbased motion capture (MoCap) framework. It can accurately capture complex human motions but this approach has some disadvantages. It require the optical sensors to be attached on the human and also demand the need of multiple camera settings. Whereas, the vision based method make use of RGB or depth image. It does not require the user to carry any devices or to attach any sensors on the human. Therefore, this methodology is getting more consideration nowadays, consequently making the HAR framework simple and easy to be deployed in many applications. Most of the vision-based HAR systems proposed in the literature used traditional machine learning algorithms for activity recognition. However, traditional machine learning methods have been outperformed by deep learning methods in recent time [5]. The most common type of deep learning method is Convolutional Neural Network (CNN). CNN are largely applied in areas related to computer vision. It consists series of convolution layers through which images are passed for processing. In this paper, we use CNN to recognise human activities from Wiezmann Dataset. We first extracted the frames for each activities from the videos. Specifically, we use transfer learning to get deep image features and trained machine learning classifiers. We applied 3 different CNN models to classify activities and compared our results with the existing works on the same dataset. In summary, the main contributions of our work are as follows: 1) We applied three different CNN models to classify human recognition activities and we showed the accuracy of 96.95% using VGG-16. 2) We used transfer learning to leverage the knowledge gained from large-scale dataset such as ImageNet [6] to the human activity recognition dataset.

II. HUMAN ACTIVITY RECOGNITION Human activity recognition is an ability to interpret human body gesture or motion via sensors and determine human activity or action [6]. Most of the human daily tasks can be simplified or automated if they can be recognized via HAR system [7], [8]. Typically, HAR system can be either supervised or unsupervised [9]. A supervised HAR system required some prior training with dedicated datasets while unsupervised HAR system is being configured with a set of rules during development. HAR is considered as an important component in various scientific research contexts i.e. surveillance, healthcare and human computer interaction (HCI) [5], [10]–[12].

A. Surveillance System In surveillance context, HAR was adopted in surveillance systems installed at public places i.e. banks or airports [7], [13], [14]. Ryoo [12] introduced a new paradigm of human activity prediction to prevent crimes and dangerous activities from occurring at public places. The findings confirmed the proposed approaches are able to recognize ongoing humanhuman interactions at the earlier stage. Lasecki et al. [15] proposed Legion:AR, a system that provides robust, deployable activity recognition by supplementing existing recognition systems with on-demand, real-time activity identification using inputs from the crowds at public places.

B. Healthcare From most of the literature reviewed, HAR is employed in healthcare systems installed in residential environment, hospitals and rehabilitation centers. HAR is used widely for monitoring the activities of elderly people staying in rehabilitation centers for chronic disease management and disease prevention [16]. HAR is also integrated into smart homes for tracking the elderly people's daily activities [17], [18]. Besides, HAR is used to encourage physical exercises in rehabilitation centers for children with motor disabilities [19], post-stroke motor patients [20], patients with dysfunction and psychomotor slowing [21], and exergaming [22]. Other than that, the HAR is adopted in monitoring patients at home such as estimation of energy expenditure to aid in obesity prevention and treatment [23] and lifelogging [24]. HAR is also applied in monitoring other behaviours such as stereotypical motion conditions in children with Autism Spectrum Disorders (ASD) at home [25], abnormal conditions for cardiac patients [26] and detection for early signs of illness [27] and it provided the clinicians with opportunities for intervention. Other healthcare related HAR such as fall detection and intervention for elderly people using HAR are found in [28]–[30].

C. Human Computer Interaction In the field of human computer interaction, HAR has been applied quite commonly in gaming and exergaming such as Kinect [31]–[33], Nitendo Wii [34], [35], full-body motionbased games for older adults [36] and adults with neurological injury [37]. Through HAR, human body gestures are recognized to instruct the machine to complete dedicated tasks. Elderly people and adults with neurological injury can perform a simple gesture to interact with games and exergames easily. HAR also enables surgeons to have intangible control of the intraoperative image monitor by using standardized free-hand movements

III. SENSING TECHNOLOGIES Generally, the sensor(s) in a conventional HAR plays an important role in recognizing human activity. Figure 1 illustrates the process of how a human activity is recognized when a body gesture is given as input. The sensor(s) capture the information acquired from human body gesture and the recognition engine analyzes the information and determines the type of activity has been performed

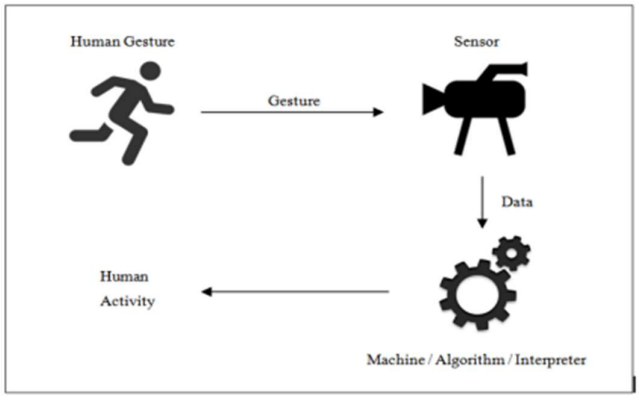

Fig: General structure of HAR system

We reviewed 32 papers published recently (from 2011 to 2014) on different sensing technologies used in HAR. These technologies are classified as RGB camera-based, depth

sensor-based and wearable-based as shown in Table I. Recognizing human activity using RGB camera is simple but with low efficiency. A RGB camera is usually attached to the environment and the HAR system will process image sequences captured with the camera. Most of the conventional HAR systems using this sensing technology are built with two major components which is the feature extraction and classification [13], [48]. Besides, most of the RGB-HAR systems are considered as supervised system where trainings are usually needed prior to actual use. Image sequences and names of human activities are fed into the system during training stage. Real time captured image sequence are passed to the system for analysis and classification by dedicated computational/classification algorithms such as Support Vector Machine (SVM) [2]. The depth sensor also known as infrared sensor or infrared camera [49] is adopted into HAR systems for recognizing human activities. In a nutshell, the depth sensor projects infrared beams into the scene and recapture them using its infrared sensor to calculate and measure the depth or distance for each beam from the sensor. The reviews found that Microsoft Kinect sensor is commonly adopted as depth sensor in HAR [33]. Since the Kinect sensor has the capability to detect 20 human body joints with its real-world coordinate [40], many researchers utilized the coordinates for human activity classification. HAR using wearable-based requires single or multiple sensors to be attached to the human body. Most commonly used sensor includes 3D-axial accelerometer, magnetometer, gyroscope and RFID tag [44], [45]. With the advancement of current smart phone technologies, many works uses mobile phone as sensing devices because most smart phones are equipped with accelerometer, magnetometer and gyroscope [29], [50]. A physical human activity can be identify easily through analysing the data generated from various wearable sensing after being process and determine by classification algorithm. Further to this, Kantoch and Augustyniak claims that GPS and temperature signal acquired from smart phone can be further feed into machine for healthcare monitoring purpose [26]

CONCLUSIONS A review has been completed on thirty two papers published in 2011-2014 for various sensing technologies used in HAR. We classify these technologies into three main categories namely RGB camera, depth sensor and wearable device. Our review found that the popularity of RGB camera in HAR research has dropped while both depth and wearable sensors are the substitutes. On the other hand, the use of Kinect sensor (depth sensor) into HAR system is promising. This could be a sign of the rise of Kinect as a popular sensing tool in HAR system.

REFERENCES

[1] A. Iosifidis, A. Tefas, and I. Pitas, “Multi-view action recognition based on action volumes, fuzzy distances and cluster discriminant analysis,” Signal Processing, vol. 93, no. 6, pp. 1445–1457, Jun. 2013.

[2] D. Weinland, R. Ronfard, and E. Boyer, “A survey of vision-based methods for action representation, segmentation and recognition,” Comput. Vis. Image Underst., vol. 115, no. 2, pp. 224–241, Feb. 2011.

[3] S. Ali and M. Shah, “Human action recognition in videos using kinematic features and multiple instance learning.,” IEEE Trans. Pattern Anal. Mach. Intell., vol. 32, no. 2, pp. 288–303, Feb. 2010.

[4] A. Oikonomopoulos and M. Pantic, “Human Activity Recognition Using Hierarchically-Mined Feature Constellations,” pp. 150–159, 2013.
[5] M. Javan Roshtkhari and M. D. Levine, “Human activity recognition in videos using a single example,” Image Vis. Comput., vol. 31, no. 11, pp. 864–876, Nov. 2013.
[6] J. Yang, J. Lee, and J. Choi, “Activity Recognition Based on RFID Object Usage for Smart Mobile Devices,” J. Comput. Sci. Technol., vol. 26, no. 2, pp. 239–246, Mar. 2011.
[7] L. Chen, H. Wei, and J. Ferryman, “A survey of human motion analysis using depth imagery,” Pattern Recognit. Lett., vol. 34, no. 15, pp. 1995–2006, Nov. 2013.
[8] W. Ong, L. Palafox, and T. Koseki, “Investigation of Feature Extraction for Unsupervised Learning in Human Activity Detection,” Bull. Networking, Comput. Syst. Softw., vol. 2, no. 1, pp. 30–35, 2013.
[9] O. D. Lara and M. a. Labrador, “A Survey on Human Activity Recognition using Wearable Sensors,” IEEE Commun. Surv. Tutorials, vol. 15, no. 3, pp. 1192–1209, Jan. 2013.
[10] A. A. Chaaraoui, J. R. Padilla-López, P. Climent-Pérez, and F. Flórez-Revuelta, “Evolutionary joint selection to improve human action recognition with RGB-D devices,” Expert Syst. Appl., vol. 41, no. 3, pp. 786–794, Feb. 2014