International Information and Engineering Technology Association
Advancing the World of Information and Engineering

Revue d'Intelligence Artificielle
Vol. 34, No. 1, February, 2020, pp. 75-80
Journal homepage: http://iieta.org/journals/ria

Human Activity Recognition Algorithm Based on One-Dimensional Convolutional Neural Network

Dengping Tang[1], Miao Jin[2*], Quan Wang[2], Wei Zhou[2], Jun Zhang[2]

[1] State Grid Hubei Electric Power Company Measurement Centre, Wuhan 430074, China
[2] China Electric Power Research Institute, Wuhan 430074, China

Corresponding Author Email: jinmiao@epri.sgcc.com.cn

https://doi.org/10.18280/ria.340110

Received: 17 July 2019
Accepted: 28 October 2019

Keywords:
human activity recognition, machine learning, convolutional neural network, sensor information

ABSTRACT

Human activity recognition (HAR) is widely used in healthcare, personal fitness, physical training and military, etc. How to distinguish various human activities accurately (such as running, walking, walking upstairs and downstairs, jumping and standing) has become an important problem in human-computer interaction. The computer vision method requires a large amount of computing resources, and it is not highly accuracy and can be easily disturbed by other objects in the background. The sensor-based method can achieve high accuracy, and it requires few computing resources, and is not disturbed by the background. This paper proposes a method based on the one-dimensional convolutional neural network (1D-CNN) to classify the sensor signals of some different activities. For comparison, this paper applies some widely used methods to accomplish the recognition task with the same dataset. Then, it tests the proposed 1D-CNN model with different datasets, for the purpose of testing its generality across users. The experimental results show that the proposed model achieves an accuracy of 95.12% with the said datasets, which is higher than those of the other methods by about 8% on average. This indicates that the proposed method has good performance in terms of generality across users, and at the same time provides a higher accuracy. The obtained results can improve the accuracy of current technologies.

1. INTRODUCTION

With the development of technologies, HAR is playing an increasingly important role in human-computer interaction. For example, in healthcare, it helps doctors observe whether a patient’s behaviour will affect his or her condition, and in personal fitness, it helps people observe their movements in real time and adjust their fitness methods. Nowadays, HAR has already become an indispensable part of our daily life. The difficulty about HAR lies in accuracy and robustness. Regarding this, much work has been carried out in the past few years, and the existing methods can be classified into traditional machine learning and deep learning.

Traditional machine learning methods usually require extracting features first, and then distinguishing various activities by using a classifier [1]. Previous work tried to extract effective movement features from the signals of tri-axis accelerometers and gyroscopes [2]. Reyes-Ortiz et al. [3] proposed a method named Transition-Aware Human Activity Recognition (TAHAR) to classify physical activities. Shoaib et al. [4] proposed three motion sensors (accelerometer, gyroscope and linear acceleration sensor) and three classifiers were used for classifying several user activities and environments. This shows that traditional machine learning methods take a cumbersome process before classification.

On the other hand, deep learning methods can directly learn the features of input data [5]. Ignatov [6] proposed using the convolutional neural network (CNN) to extract local features as well as simple statistical features that preserve information about the global form of time series. The research showed that the deep learning method can achieve improved performance compared with traditional machine learning methods in practical applications [7]. It not only can save a lot of work time, but also achieve high accuracy.

However, the existing methods are still not sufficiently accurate - they cannot achieve the same accuracy when used on self-created datasets. In order to improve the accuracy on self-created datasets, this paper proposes a CNN-based method to classify six different activities. This method achieves an average accuracy of 95.12%, which is about 8% higher than those of other methods. It shows that the proposed method has good generality across users while maintaining a higher accuracy.

The rest of this paper is organized as follows: Section II describes the research on HAR; Section III introduces the proposed CNN method for HAR; Section IV provides the verification of the proposed CNN method and compares it with other widely used methods; and Section V gives the conclusion and some ideas about future works.

2. RELATED WORK

Human activity recognition has long been a research hot spot. So far, many solutions to HAR have been proposed. To sum up, these attempts can be classified into traditional machine learning methods and deep learning methods.

2.1 Traditional machine learning method

Traditional machine learning was widely applied in the early stage of research on this topic. This kind of method usually requires extracting features first.

The key processing stages mainly include data pre-processing, object segmentation, feature extraction, and classifier implementation [8]. Trost et al. [9] used sensors worn on the wrist and the hip to detect seven physical activities and used logistic regression as the classifier. This research showed the potential of using the wrist position for activity recognition, but it only evaluated these two positions separately without combining them.

Previous work tried to extract effective body movement features from the signals of tri-axis accelerometers and gyroscopes [4]. Van Nguyen et al. reduced the dimensionality and size of data, and built a recognition model corresponding to each feature extraction method, and also used the Support Vector Machine (SVM) model for training. Shen et al. [1] analyzed the extraction of time, frequency and wavelet-domain features from the acceleration data of motion sensors, and used several methods to recognize six behaviours, such as SVM, KNN, MLP and Random Forest. The highest accuracy rate is 90.65%. Chen and Shen [10] used a similar method and reported an accuracy of 95%. Janidarmian et al. [11] conducted an extensive analysis on feature representations and classification techniques of activity recognition. It performed principal component analysis to reduce feature vector dimensions while keeping essential information. Shoaib et al. [4] used three motion sensors and three classifiers, evaluated the effect of seven window sizes on thirteen activities and showed how increasing the window size would affect these various activities in different ways. Kang and Han [12] proposed a novel step detection algorithm that detects each step by peak value with some pseudo peak points ignored. Shoaib et al. [13] evaluated the effect of seven window sizes on thirteen activities and showed how increasing the window size would affect these various activities in different ways.

2.2 Deep learning method

Recently the more widely used method is deep learning. In this kind of method, the features of input data can be learnt directly. In addition, CNN is very useful for fully automatic extraction of discriminative features from raw sensor data [14].

Early work on the use of deep learning methods for HAR was based on deep belief network (DBN), which was built by stacking multiple layers of restricted Boltzmann machines (RBM). Subsequent DBN-based models exploited the intrinsic temporal sequences in human activities by implementing the hidden Markov model (HMM) above the RBM layers [15]. They performed an unsupervised pre-training step to generate intrinsic features and then used the available data labels to tune the model. However, HMMs are limited by their numbers of possible hidden states and become impractical when modeling long-range dependencies in large context windows.

For the dataset with one-dimensional time series, Ignatov [6] used the CNN to extract local features and simple statistical features. Chen and Xue [5] proposed a CNN-based method to identify eight activities, for which there is one input channel and three layers, and reported an accuracy of 93.8%. Ronao and Cho [16] proposed a deep convolutional neural network, which performs efficient and effective HAR by exploiting the inherent features of activities and 1D time-series signals. Murad and Pyun [17] proposed the use of the deep recurrent neural network (DRNN) for building recognition models that are capable of capturing long-range dependencies in variable-length input sequences. Shahroudy et al. [18] introduced a large-scale dataset for RGB+D human action recognition with more than 56 thousand video samples and 4 million frames, collected from 40 distinct subjects, and proposed a new recurrent neural network structure to model the long-term temporal correlation of the features for each body part, and also utilized them for better action classification. Singh et al. [19] demonstrated the use of the CNN and a comparison of results, which has been performed with Long Short Term Memory (LSTM). Xue et al. [20] realized the visualization of the sensor-based activity's data features extracted from the neural network and sent the features to the DNN-based fusion model, with a reported accuracy of 96.1%.

However, the results were not satisfactory when our self-created datasets are tested by the existing methods. In order to get a higher accuracy, this paper proposes a 1D-CNN method. The experiment results show that the proposed method can achieve good results on different datasets, indicating that the proposed model has good generality across users and also has a higher accuracy.

3. ONE-DIMENSIONAL CONVOLUTIONAL NEURAL NETWORKS

Recently, the deep neural network architecture has seen significant improvements in many fields of pattern recognition. This section introduces the proposed 1D-CNN method.

3.1 Data collection

In this evaluation, a microcontroller and an accelerometer are used to collect acceleration data in real time. The sensor is calibrated in such a way that the accelerations in 3 axes are zero when it is placed on a flat and horizontal surface [9]. The motion sensor is fixed on a person's waist, and the x-axis, y-axis and z-axis acceleration data reflect the front side, right side and lower side of the person, respectively. Motion features collected include walking, running, standing, walking upstairs, walking downstairs and jumping.

As the original acceleration data cannot be directly used as training datasets, they are divided into short pieces with the help of overlapped time windows. The length of each time window is set to 2 second, and the degree of overlapping is 50%.

3.2 Model training

First let us see what is activation function. In an artificial neural network, the activation function of a node defines the output of that node given an input or set of inputs.

The rectifier is an activation function in the artificial neural network. Compared with traditional neural network activation functions, such as logistic sigmoid and hyperbolic function, it is faster and more effective. In this paper, the Rectified Linear Unit (ReLU) is used as the activation function.

$$y = \max(0, w^{T}x + b) \tag{1}$$

Eq. (1) is the activation function used in this paper, where x is the vector from the last layer of the neural network, and y is the input to the next layer of the neural network.

Then, it is the introduction of the convolutional layer. In image processing, convolution is the first layer of the features extraction from an input image. Convolution learns image features using small squares of input data to preserve the relationships between pixels.

The biggest distinction between acceleration data and image data lies in the size - the size of image data is two-dimensional while that of acceleration data is one-dimensional. Therefore, the convolution kernel width is adjusted to 1.

$$S(n) = (f \cdot g)[n] = \sum_{m=0}^{N-1} f(m)g(n-m) \tag{2}$$

The convolutional layer is shown as (2), where $S(n)$ is the convolution result, and N is the length of the signal $f(m)$.

The next is about pooling layer. If the images are too large, the pooling layer will reduce the number of parameters. Spatial pooling reduces the dimensionality of each image mapped but still retains important information. This is partially done by providing an abstracted form of representation to help over-fitting. Also, it reduces the computational cost by reducing the number of the parameters to learn and provides basic translation invariance for the internal representation.

This paper uses the Max pooling, which takes the largest element from the rectified feature map.

The model in Chen and Xue [5] is a 3-layer 2D-CNN model with 1 input channel. Considering that there are no stationary properties among three acceleration values for each sample, the input channel number is adjusted to 3 and the 1D-CNN is defined.

The detail of the CNN model is shown in Figure 1, which contains 3 convolutional layers and 3 pooling layers. The model works directly on the raw acceleration signals without any other processing.

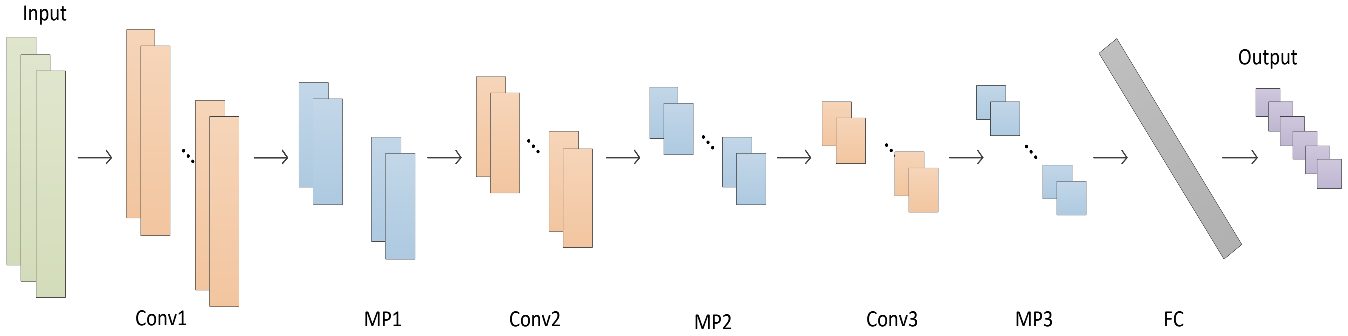


Figure 1. The proposed 1D-CNN model, where Conv, MP and FC stand for Convolution, Max Pooling, and Fully Connected respectively

The last is about the optimization algorithm in the proposed model.

Deep learning is a highly iterative process. Various permutations of the parameters should be tried to figure out which combination works best.

Gradient descent is the most common method of optimization algorithm; however, one of the disadvantages of gradient descent is that it begins parameter updating only after it goes through all training data. This poses a challenge when the training data are too big to fit in the computer memory.

Mini-batch gradient descent is a clever workaround that tackles the above problem of gradient descent. It finds a balance between the robustness of stochastic gradient descent and the efficiency of batch gradient descent, and that is why it is used in this paper.

$$W = W - \alpha\nabla J(W, b, x^{(z:z+\delta)}, y^{(z:z+\delta)}) \tag{3}$$

In mini-batch gradient descent, the cost function is averaged over a small number of samples, which looks like (3), where δ is the mini-batch size.

4. EXPERIMENT AND ANALYSIS

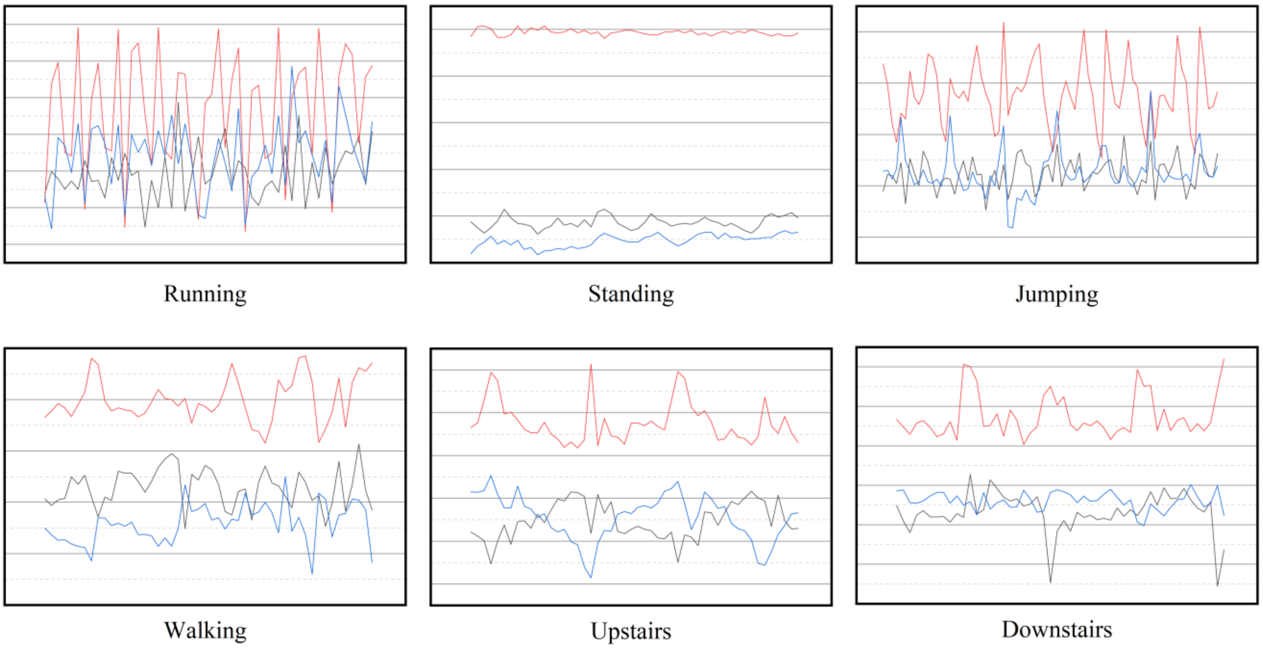
This section first introduces the datasets collection. Next, it describes the details of the experiments. Finally, the proposed model is compared with other general models and a conclusion is drawn.

4.1 Datasets

In this evaluation, a microcontroller and an accelerometer are used to collect acceleration data in real time. The model of the microcontroller is Arduino UNO, and the model of accelerometer is ADXL345. The sensor is calibrated in such a way that the accelerations in 3 axes are zero when it is placed on a flat and horizontal surface. The motion sensor is fixed on the person’s waist, and the x-axis, y-axis and z-axis acceleration data reflect the front side, right side and lower side of the person, respectively.

The sampling rate is 75Hz. Then the acceleration data are sent to the server by the ZigBee module. In order to maintain the consistency of data, the sensor needs to be fixed on the waist.

Six motions are selected, including walking, running, standing, walking upstairs, walking downstairs and jumping. With the aim of training a generalized model, 1100000 motion sequences of six different motions are selected from 100 healthy subjects (50 males and 50 females). The proportions of walking, running, standing, walking upstairs, walking downstairs and jumping are 37%, 30%, 6%, 11%, 10% and 6%. Figure 2 shows the samples of the dataset.


Running

Standing

Jumping

Walking

Upstairs

Downstairs

Figure 2. Sample of our dataset

4.2 Pre-processing

Since the collected data are continuous and long-lasting, they cannot be applied directly as the training or recognition datasets. Therefore, the acceleration data are divided into short pieces with the help of the overlapped time windows. The length of each time window is set to 2 second, and the degree of overlapping is 50%. The motion sequences are then normalized. The acceleration data are cropped into a matrix with a size of 150*3.

After simple pre-processing, 1100000 motion sequences are obtained. They are divided randomly into two subsets, 70% of which are for training and 30% for testing.

4.3 Experimental results

Motion representation is extracted for a 2-second time segment. At every second, a 2-second time segment is processed, which overlaps with the previous processed time segment by 1 second. To carry out the training from scratch, the base learning rate is set to 0.001, the step size 1000, the momentum 0.9 and the weight decay 0.0005. During fine tuning, the base learning rate is adjusted to 0.001.

The first is to test whether the proposed 1D-CNN model is reliable or not. Here some other CNN models are used for comparison of performance, such as the 3-layer 2D-CNN 0, the 1-layer 1D-CNN, and the 2-layer 1D-CNN. The experimental results are shown in Figure 3, the confusion matrix of the proposed model is shown in Figure 4 and the precision and recall of the proposed model are shown in Figure 5.

Then, the proposed 1D-CNN model is tested on different datasets and compared with other methods to find out its generality across users. The experimental results are shown in Figure 6. The Actitracker Dataset [21] contains six motions, namely walking, running [22], walking upstairs, walking downstairs, jumping and standing. The Hand Motion Dataset [23] contains six hand motions, namely lifting, picking up, putting down, pulling, staying put and walking. This dataset is collected from the right hands of eight users.

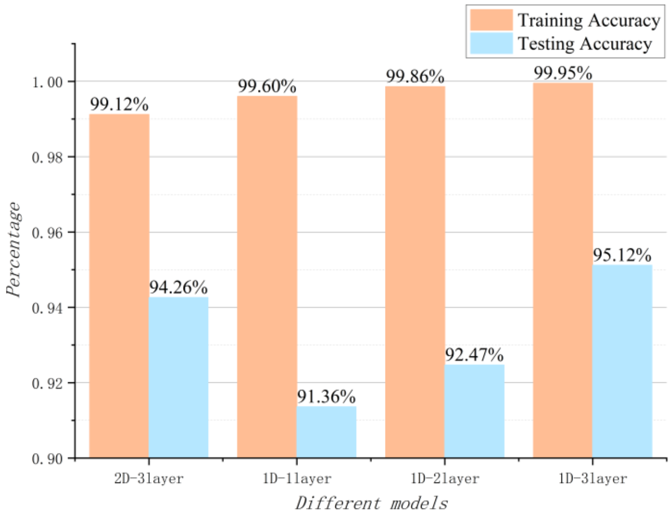
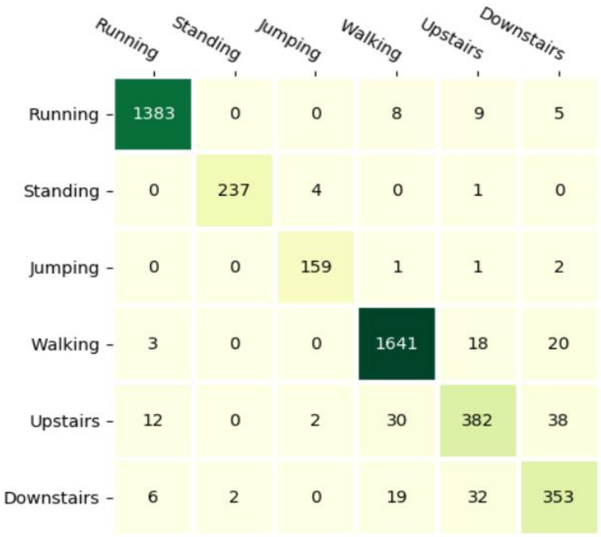


Figure 3. Accuracy of different models



	Running	Standing	Jumping	Walking	Upstairs	Downstairs
Running	1383	0	0	8	9	5
Standing	0	237	4	0	1	0
Jumping	0	0	159	1	1	2
Walking	3	0	0	1641	18	20
Upstairs	12	0	2	30	382	38
Downstairs	6	2	0	19	32	353

Figure 4. Confusion matrix of the proposed CNN model

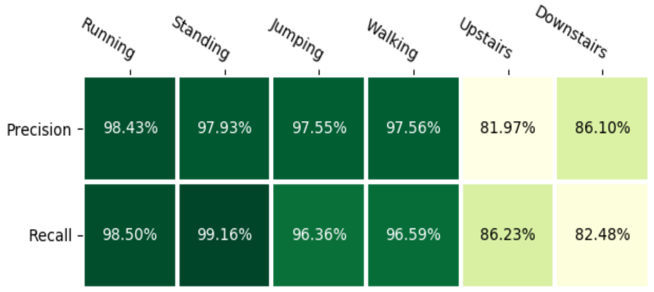
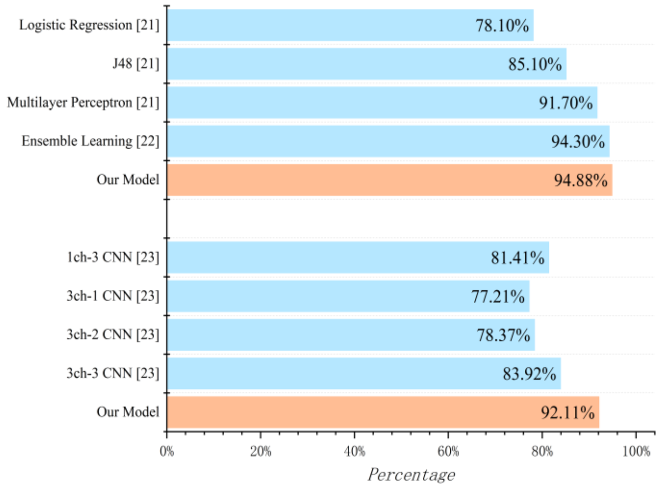


Figure 5. Precision and recall of the proposed CNN model



*The first dataset in figure is Actitracker Dataset [21]. The second dataset in figure is Hand Motion Dataset [23].

Figure 6. Accuracy of the proposed model on different datasets and comparison with other methods

4.4 Analysis

The experimental results in Figure 3 show the testing accuracy of the 3-layer 2D-CNN is higher than those of the 1-layer 1D-CNN and the 2-layer 1D-CNN, but the testing accuracy of the 3-layer 2D-CNN is lower than that of the 3-layer 1D-CNN. The proposed model works well, with an accuracy of 95.12%. All of this shows that the 3-layer 1D-CNN is more suitable for our HAR dataset, and also proves that the proposed model is effective for a broad range of activity recognition tasks.

From Figure 5, it can be seen that the accuracy of walking upstairs and walking downstairs is relatively low, but the proposed model works well for running, standing, jumping and walking.

The experimental results in Figure 6 show that the accuracy rate is very stable with respect to different datasets. In the Actitracker Dataset, the proposed 1D-CNN model achieves an accuracy of 94.88%, and in the Hand Motion Dataset, the proposed 1D-CNN model achieves an accuracy of 92.11%. The accuracy is improved by 11.88% compared with those of the previous methods. This indicates that the proposed method has good generality across users.

According to the experimental results, the proposed model has the highest accuracy. The performance results clearly demonstrate that the 1D-CNN is very effective for HAR.

The main reason why the performance of the proposed model is good for HAR tasks is that it contains sufficient deep layers, enabling the model to extract effective discriminative features. These features are exploited to classify activities and even further applied to perform more complex behaviour recognitions tasks.

5. CONCLUSION

This paper proposes an acceleration-based HAR algorithm using CNN. Additionally, it empirically evaluates the proposed model through experiments. According to the characteristic of the acceleration data, this paper modifies the conventional CNN structure. Then it designs experiments to evaluate the recognition performance of the proposed 1D-CNN and other widely used methods. The experiments are executed on a large dataset of six typical kinds of activities, which has 1100000 samples from 100 subjects. The experimental results show that the proposed 1D-CNN model works well, with an accuracy of 95.12%. From the generality testing, it can be seen that the proposed method has good generality across users and also has a higher accuracy. The proposed 1D-CNN model is accurate and robust without any feature extraction, and it is also suitable for building a real-time HAR system on a mobile platform.

In this paper, the data collection is performed in a simple environment. However, in real-life scenarios, these activities can be performed in various ways. The recognition of such complex activities is yet to be explored further. This research is an attempt towards this direction, and in the future, we plan to conduct more data collection experiments in real-life setups.

ACKNOWLEDGMENT

This work was supported by STATE GRID Corporation of China, the science and technology projects of research on key technologies for safety management of metrological operations based on human activity recognition.

REFERENCES

[1] Shen, C., Chen, Y., Yang, G. (2016). On motion-sensor behaviour analysis for human-activity recognition via smartphones. 2016 IEEE International Conference on Identity, Security and Behaviour Analysis (ISBA). https://doi.org/10.1109/isba.2016.7477231

[2] Jalal, A., Kim, Y.H., Kim, Y.J., Kamal, S., Kim, D. (2017). Robust human activity recognition from depth video using spatiotemporal multi-fused features. Pattern Recognition, 61: 295-308. https://doi.org/10.1016/j.patcog.2016.08.003

[3] Reyes-Ortiz, J.L., Oneto, L., Samà, A., Parra, X., Anguita, D. (2016). Transition-aware human activity recognition using smartphones. Neurocomputing, 171: 754-767. https://doi.org/10.1016/j.neucom.2015.07.085

[4] Shoaib, M., Bosch, S., Incel, O.D., Scholten, H., Havinga, P.J. (2016). Complex human activity recognition using smartphone and wrist-worn motion sensors. Sensors, 16(4): 426. https://doi.org/10.3390/s16040426

[5] Chen, Y., Xue, Y. (2015). A deep learning approach to human activity recognition based on single accelerometer. 2015 IEEE International Conference on Systems, Man, and Cybernetics, 2015: 1488-1492. https://doi.org/10.1109/smc.2015.263

[6] Ignatov, A. (2018). Real-time human activity recognition from accelerometer data using Convolutional Neural Networks. Applied Soft Computing, 62: 915-922. https://doi.org/10.1016/j.asoc.2017.09.027

[7] Ravi, D., Wong, C., Lo, B., Yang, G.Z. (2016). Deep learning for human activity recognition: A resource efficient implementation on low-power devices. IEEE International Conference on Wearable & Implantable Body Sensor Networks. https://doi.org/10.1109/bsn.2016.7516235

[8] Subasi, A., Dammas, D.H., Alghamdi, R.D., Makawi, R.A., Albiety, E.A., Brahimi, T., Sarirete, A. (2018). Sensor based human activity recognition using adaboost ensemble classifier. Procedia Computer Science, 140: 104-111. https://doi.org/10.1016/j.procs.2018.10.298

[9] Trost, S.G., Zheng, Y., Wong, W.K. (2014). Machine learning for activity recognition: hip versus wrist data. Physiological Measurement, 35(11): 2183-2189. https://doi.org/10.1088/0967-3334/35/11/2183

[10] Chen, Y., Shen, C. (2017). Performance analysis of smartphone-sensor behaviour for human activity recognition. IEEE Access, 5: 3095-3110. https://doi.org/10.1109/access.2017.2676168

[11] Janidarmian, M., Roshan Fekr, A., Radecka, K., Zilic, Z. (2017). A comprehensive analysis on wearable acceleration sensors in human activity recognition. Sensors, 17(3): 529. https://doi.org/10.3390/s17030529

[12] Kang, W., Han, Y. (2015). SmartPDR: Smartphone-based pedestrian dead reckoning for indoor localization. IEEE Sensors Journal, 15(5): 2906-2916. https://doi.org/10.1109/jsen.2014.2382568

[13] Shoaib, M., Bosch, S., Incel, O.D., Scholten, H., Havinga, P.J. (2016). Complex human activity recognition using smartphone and wrist-worn motion sensors. Sensors, 16(4): 426. https://doi.org/10.3390/s16040426

[14] Huang, J., Lin, S., Wang, N., Dai, G., Xie, Y., Zhou, J. (2019). TSE-CNN: A two-stage end-to-end CNN for human activity recognition. IEEE Journal of Biomedical and Health Informatics. https://doi.org/10.1109/jbhi.2019.2909688

[15] Alsheikh, M.A., Selim, A., Niyato, D., Doyle, L., Lin, S., Tan, H.P. (2016). Deep activity recognition models with triaxial accelerometers. Workshops at the Thirtieth AAAI Conference on Artificial Intelligence.

[16] Ronao, C.A., Cho, S.B. (2016). Human activity recognition with smartphone sensors using deep learning neural networks. Expert Systems with Applications, 59: 235-244. https://doi.org/10.1016/j.eswa.2016.04.032

[17] Murad, A., Pyun, J.Y. (2017). Deep recurrent neural networks for human activity recognition. Sensors, 17(11): 2556. https://doi.org/10.3390/s17112556

[18] Shahroudy, A., Liu, J., Ng, T.T., Wang, G. (2016). Ntu rgb+ d: A large scale dataset for 3d human activity analysis. Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp. 1010-1019. https://doi.org/10.1109/cvpr.2016.115

[19] Singh, D., Merdivan, E., Hanke, S., Kropf, J., Geist, M., Holzinger, A. (2017). Convolutional and recurrent neural networks for activity recognition in smart environment. Towards Integrative Machine Learning and Knowledge Extraction, 194-205. https://doi.org/10.1007/978-3-319-69775-8_12

[20] Xue, L., Xiandong, S., Lanshun, N., Jiazhen, L., Renjie, D., Dechen, Z., Dianhui, C. (2018). Understanding and improving deep neural network for activity recognition. arXiv preprint arXiv:1805.07020. https://doi.org/10.4108/eai.21-6-2018.2276632

[21] Kwapisz, J.R., Weiss, G.M., Moore, S.A. (2011). Activity recognition using cell phone accelerometers. ACM SIGKDD Explorations Newsletter, 12(2): 74-82. https://doi.org/10.1145/1964897.1964918

[22] Catal, C., Tufekci, S., Pirmit, E., Kocabag, G. (2015). On the use of ensemble of classifiers for accelerometer-based activity recognition. Applied Soft Computing, 37: 1018-1022. https://doi.org/10.1016/j.asoc.2015.01.025

[23] Wu, T.Y., Chien, T.A., Chan, C.S., Hu, C.W., Sun, M. (2017). Anticipating daily intention using on-wrist motion triggered sensing. Proceedings of the IEEE International Conference on Computer Vision, pp. 48-56. https://doi.org/10.1109/iccv.2017.15