Signal, Image and Video Processing (2021) 15:1739–1746
https://doi.org/10.1007/s11760-021-01907-4

ORIGINAL PAPER

Enhanced bag-of-words representation for human activity recognition using mobile sensor data

Rasel Ahmed Bhuiyan[1] · Shams Tarek[1] · Hongda Tian[2]

Received: 19 July 2020 / Revised: 19 February 2021 / Accepted: 1 April 2021 / Published online: 27 April 2021

Abstract

Human activity recognition based on sensors (e.g., accelerometer and gyroscope) embedded in smartphones is of great significance for many applications under uncontrolled environments. Although significant progress has been noticed in this field, one of the challenges limiting its real-life applications lies in robust feature extraction for efficient activity recognition on smartphones. This study addresses this challenge by proposing an improved bag-of-words representation for activity signal characterization. Specifically, raw activity signals are processed by discrete wavelet transformation to extract local features, which will be clustered using K-means to form a bag-of-words dictionary. The vocabularies in the dictionary are regarded as bin centers for histogram feature construction. For each local feature of an activity signal, its distance from all the bin centers will be measured. To capture higher-order information for feature representation, the frequency for the bin centers corresponding to the minimum *n* distances will be updated. Moreover, the frequency is increased by a trigonometry constraint cosine value of the corresponding distances to account for activity signals' structural information. The proposed feature representation has been verified with three well-established classifiers, namely SVM, ANN, and KNN on the UCI-HAR dataset. The consistently good performance validates the effectiveness and robustness of the proposed feature representation. Compared with the state-of-the-art, the experimental results also demonstrate the advantage of the proposed method in terms of accuracy and computational cost.

Keywords Human activity recognition · Bag-of-words · Histogram representation · Higher-order information

1 Introduction

Human activity recognition (HAR) has become a popular research topic in the field of ubiquitous computing and human–computer interaction [1–3]. It plays a crucial role in a wide range of real-life applications such as health care, elder care, smart home, fitness tracking, sports tracking, robotics, social security, and so on [4–9].

✉ Rasel Ahmed Bhuiyan
raselcse34@gmail.com

Shams Tarek
shamstarekargho@gmail.com

Hongda Tian
Hongda.Tian@uts.edu.au

1 Department of Computer Science and Engineering, School of Engineering, Uttara University, Dhaka, Bangladesh

2 Data Science Institute, Faculty of Engineering and Information Technology, University of Technology Sydney, Sydney, Australia

Although contextual information captured by cameras plays an important role in activity recognition [10–12], video-based HAR is only suitable for applications under controlled environments. In contrast, HAR using body-worn sensors is more applicable under uncontrolled indoor and outdoor environments. It involves collecting information from a set of body-worn motion sensors placed at different body locations such as chest, wrist, and ankle [13]. However, this kind of HAR is less user friendly due to the fact that these sensors might cause inconvenience to the users. Moreover, data collection cannot be done surreptitiously. Nowadays, smartphones have become a critical part in our day-to-day lives, which carry built-in inertial sensors such as accelerometer and gyroscope. The data collected by these embedded sensors could be exploited for activity recognition as long as the user carries a smartphone while performing activities [14]. The major advantage of this type of HAR is that the data can be collected unobtrusively without the help of any additional hardware [12]. Furthermore, it is appropriate for analyzing user's daily activities continuously. In this study, we will focus on HAR using gyroscope and accelerometer data collected by smartphones.

HAR based on mobile sensor data usually follows a four-step approach including data segmentation, feature extraction, feature selection, and classification [12]. Among all these steps, feature extraction is the most important step because of its impact on the performance of classification models [15]. Feature extraction from sensor data has always been a challenging task due to the facts such as variable orientation of smartphones, placement and subject, unification of simple activities leading to a complex activity, which will degrade the recognition accuracy to a great extent [2,6]. One approach that is widely used is to extract various hand-crafted features such as spectral entropy, energy of different frequency bands, auto-regressive and FFT coefficients [16,17]. Though this approach often shows good performance in practice, it relies on domain-specific knowledge and its generalization to new data sources is usually mediocre [18]. As another alternative deep learning-based algorithms have been proposed in HAR systems to make the feature extraction process automatic [1,19,20]. In addition to high accuracy and good generalization, one key advantage of this approach is that training of a deep learning model can be done in an end-to-end fashion, which is completely automated and removes the step of manual feature engineering [4,15,21–23]. However, to gain this capability, this approach requires large-scale annotated samples for performing the training task. Assembling such a vast dataset of annotated cases is time-consuming and again considered an arduous task. Carrying out the annotation on new data will also be very tiresome and expensive. Also, the computational cost is required for this approach.

This paper aims to propose a low-cost solution capable of automatically extracting quality features without domain-specific knowledge for human activity recognition. Our proposed feature representation is derived from Bag-of-Words (BoW) [24] model, which is a very popular information retrieval (IR) algorithm for text, image, and signal classification due to its simplicity and efficiency. Existing BoW-based methods mainly include two parts: (a) vector quantization using clustering, and (b) histogram representation based on minimum distance calculation and frequency value distribution. For histogram representation, the frequency value is increased by *one* for the bin center which has a minimum distance from an activity signal. Such kind of histogram representation only considers first-order information in terms of bin center selection. The frequency value update practice also neglects the temporal and structural information. All of these restrict the discriminative power of the extracted features. In this work, we first extract local features from raw activity signal using discrete wavelet transformation (DWT). The local features are then clustered using K-means to generate BoW vocabulary. For a sample activity signal, it may share similar close distances from multiple bin centers. Based on this understanding, we consider the minimum n distances instead of only *one* for bin center selection by incorporating higher-order information. Furthermore, to capture structural information of activity signals, the frequency value is updated with the cosine value of minimum distances instead of *one*. The key contributions of this paper are as follows:

1. We propose a simple and efficient feature extraction scheme for human activity recognition using mobile sensor data.
2. We enhance the BoW-based representation by incorporating higher-order structural information of activity signals.
3. We test the proposed feature representation with three widely used classifiers namely SVM, ANN, and KNN and achieve consistently good performance.
4. We compare the proposed method with other state-of-the-art and demonstrate it outperforms the baselines in terms of both effectiveness and efficiency.

Following this introduction, we review the related literature in Sect. 2. The proposed feature representation is presented in Sect. 3. The experiments are described in Sect. 4. We conclude the paper in Sect. 5.

2 Related work

Discriminative feature extraction is the key for HAR. In this regard, coordinate transformation and principal component analysis (CT-PCA) for feature extraction has been explored in [6]. A group-based context-aware approach has been initiated to resolve the issue in [25]. The dimensionality reduction-based feature extraction scheme using PCA has been proposed in [26] but this HAR system cannot reach the baseline. Still, high dimensional feature vectors sometimes create a phenomenon namely 'curse of dimensionality' and thus a hybrid feature extraction method using sequential floating forward search (SFFS) has been proposed in [27]. Feature extraction and dimension reduction model has been introduced in [28], where signal impulses are extracted using EPS and LDA is utilized for dimensionality reduction of the extracted impulses. Though these procedural machine learning models sometimes show good performance, they rely on domain-specific knowledge. To solve this problem, the deep learning-based approach becomes popular in the HAR system [22,29]. A pre-trained convolutional neural network (CNN) has been introduced in [21] to make the feature extraction automated. Time segment of data is very important for reducing computational cost. A method using logistic model tree (LMT) for HAR has been suggested in this regard [30]. A data-driven approach has also been followed in HAR [31].

Most of the classification methods lack incremental learning ability which can be achieved using probabilistic neural network and fuzzy cluster algorithm but this comes at a price of reducing the recognition accuracy [32]. To reduce the backward locking phenomenon, layer-wise CNN with local loss has been used, and to train the network with local loss, they implemented two different local learning networks to update and optimize each hidden layer, which shows promising results with a smaller number of parameters but might show some discrepancy if the number of parameters is increased [33]. In some scenarios, virtual wearable sensors are also used to recognize complex real-scene data using deep learning [34]. Stack AutoEncoder (SAE) and artificial bee colony (ABC) along with heuristic optimization algorithm are also being used in the field of HAR [19]. A new deep learning approach using computer vision which converts time series data into images is also proposed for HAR [35]. One of the greatest shortcomings of semisupervised learning is the imbalance distribution of labeled data over classes. In [20], the authors suggest a semisupervised deep model for imbalanced activity recognition. Furthermore, an unsupervised learning technique is proposed in [36]. In addition to high accuracy and good generalization, one main advantage of deep learning approach is that the model can be trained in an end-to-end fashion which completely removes the need of manual feature engineering. However, high computational cost is required for this approach. To address this issue, this paper aims to propose a low-cost solution capable of automatically extracting quality features for human activity recognition.

3 Proposed BoW-based representation

This section depicts our improved BoW-based feature representation for human activity recognition, which is a two-step procedure consisting of BoW vocabulary generation and BoW-based histogram feature extraction.

3.1 BoW vocabulary generation

3.1.1 Local feature extraction

Generally, the raw signal data of human activities collected from mobile sensors need to be converted to time-frequency features for further analysis. At high frequencies, the time resolution is decent, but frequency resolution is poor. At low frequencies, the frequency resolution is excellent, but time resolution is unsatisfactory. Fortunately, most real-life signals have high-frequency components for a short duration and low-frequency components for a long duration. DWT provides a proper trade-off between time and frequency resolution. It could help to measure both fast and slow activities. Thus, a single-level one-dimensional DWT is employed to extract local features of each axial signal of a sample activity signal. More specifically, the Haar wavelet mother function is adopted, and the approximate coefficient is calculated as local features in this study. The Haar wavelet transform is well known for its being effective, conceptually simple, memory-efficient, and fast.

3.1.2 Bin center formation

This paper suggests a scheme for histogram construction based on the extracted local features. Specifically for each class, 10% signals are randomly selected from the whole training set for BoW vocabulary formation. The local features of the selected raw signals from the same class are extracted using DWT. Then, they are fed into a K-means clustering algorithm to obtain k cluster centers, which are considered as the BoW vocabularies, $C_i = \{c_1^i, c_2^i, c_3^i, \ldots, c_k^i\}$ for the ith class. The BoW vocabularies for all classes built in a similar way are grouped together to form the whole BoW dictionary

$$B = \bigcup_{i=1}^{N} C_i \tag{1}$$

where N is the total number of considered classes. Later, in both training and test phases, the vocabularies in B are regarded as the histogram bin centers to extract a BoW-based histogram feature for each raw signal.

3.2 BoW-based histogram feature extraction

Considering a signal S having P local features, we need to construct a BoW-based histogram for activity recognition. For this purpose, a mapping of the local signal features to the nearest vocabulary has been conducted. Specifically for a local feature L_p of signal S, its distance $D_{p,q}$ from a BoW histogram bin center B_q is measured as follows:

$$D_{p,q} = ||L_p - B_q||_2, \quad \forall q \in \{1, 2, \ldots, kN\}. \tag{2}$$

Let D_p be a set including the distance values between L_p and all the vocabularies in B. We select n minimum distances from D_p through:

$$\text{argmin}_{D'_p \subset D_p, |D'_p|=n} \quad D_{p,l} \tag{3}$$

where l denotes the corresponding vocabulary indexes. By this means, we aim to capture higher-order information to account for similar close distances from multiple bin centers. For the selected bin centers, their frequency value H_l is increased by a trigonometry constraint cosine value of the

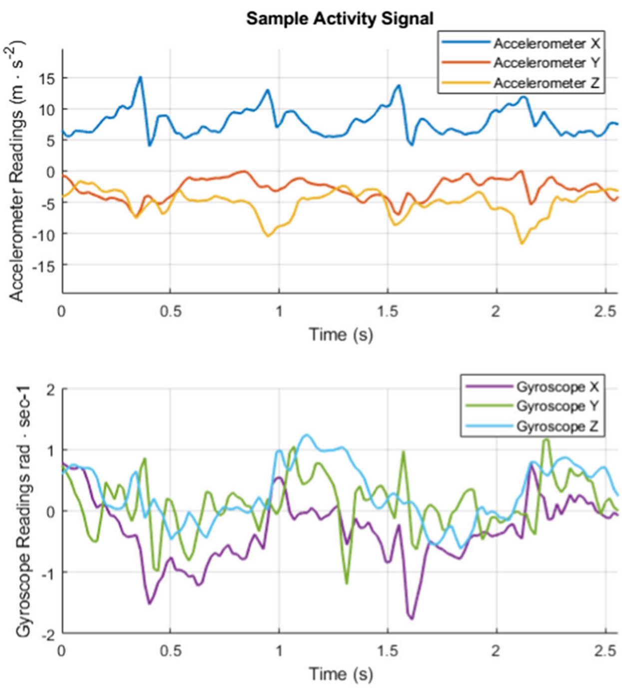

Fig. 1 Axial representation of a sample activity signal

corresponding distances as follows:

$$H_l \leftarrow H_l + \text{cosine}(D_{p,l}), \quad \forall p \in \{1, 2, \ldots, P\}. \tag{4}$$

It is expected that the structural information of activity signals could be captured in this way. A repeated process is performed for all P local features of S and a histogram is constructed. Such BoW histogram bin frequencies are used as features for activity recognition in this study.

4 Experiments

4.1 Dataset

The UCI-HAR dataset [14] was constructed with the help of 30 volunteers whose age varies from 19 to 48. They performed five activities which include walking, laying, sitting, standing, and climbing stairs (both upstairs and downstairs), while carrying a smartphone on the waist. Triaxial linear acceleration as well as triaxial angular velocity were monitored and captured at a constant rate of 50 Hz. Figure 1 shows the axial representation of a sample activity signal. With the help of low pass filters and then sampling in fixed-width sliding windows of 2.56 s with 50% overlap, both accelerometer and gyroscope signals were pre-processed. Thus, the dimension of each axial signal is 128. In order to label the data manually, the performed activities have been video-recorded. The whole dataset has been randomly divided into two sets.

Table 1 Class-wise data description of UCI-HAR dataset

Activity	Number of signals	Signal dimension
Laying	1944	768
Sitting	1777	768
Standing	1906	768
Walking	1722	768
Climbing stairs	2950	768

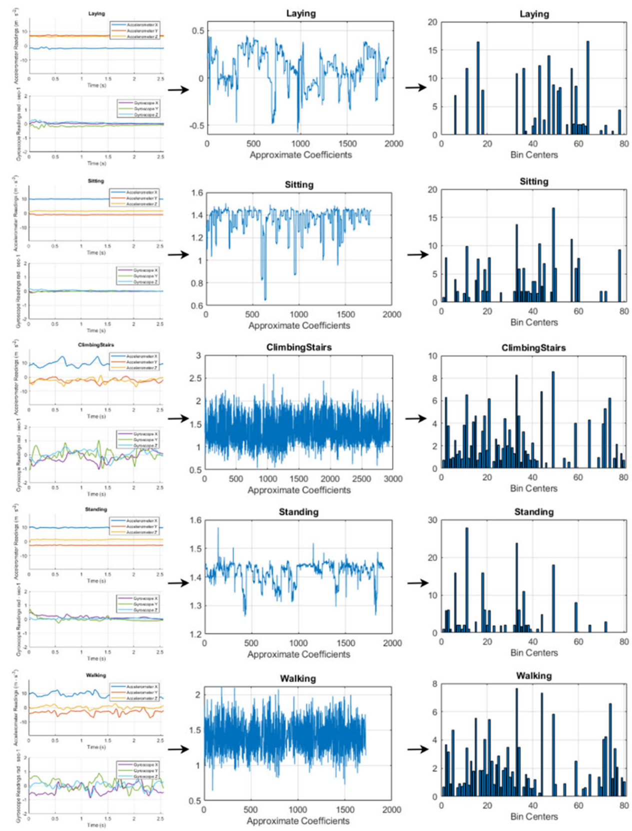

Fig. 2 BoW histograms for human activities. The leftmost column presents the signals of five activities, the middle column presents the local features, and the rightmost column presents the BoW histograms corresponding to each activity

Among all the volunteers, 70% of the data were used for generating training data and 30% for testing data. Class-wise data distribution is shown in Table 1.

4.2 Experimental setup

BoW-based features extracted using the proposed method need to be input to classifiers for human activity recognition. To verify the robustness of our proposed feature representation, three well-known classifiers, namely SVM, ANN, and KNN are all tested. The SVM classifier is implemented for multi-class classification using the 'one versus all' coding scheme with regularization parameter $C \in \{1, 2, 4, \ldots, 128\}$

Table 2 The best performance achieved by each classifier along with the optimal classifier settings

Classifier	Classifier Setting	Performance (%)			
		Accuracy	Precision	Recall	F_1 Score
SVM	$C = 32$ $\sigma = 32$	96.41	96.55	96.14	96.34
ANN	'Trainscg' function with $N = 40$	95.78	95.98	95.49	95.71
KNN	**'Cityblock' distance with $K = 3$**	**98.10**	**98.03**	**97.97**	**98.00**

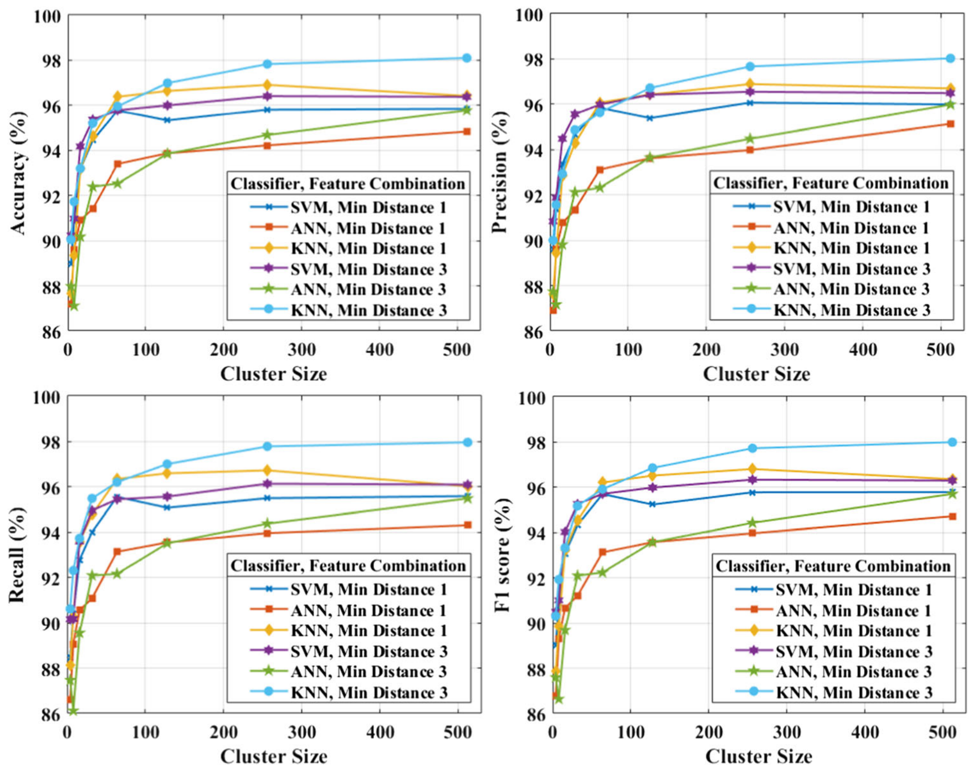

Fig. 3 Accuracy, precision, recall, and F_1 score results based on different combinations of feature extraction and classifiers

and Gaussian Radial Basis Function (RBF) kernel parameter $\sigma \in \{1, 2, 4, \ldots, 128\}$. The ANN classifier is implemented for the 'trainscg,' 'trainrp,' 'trainbfg,' 'trainlm,' and 'traingd' training functions in MATLAB and for a hidden node size of $N \in \{10, 20, 30, \ldots, 100\}$. The KNN classifier is implemented for 'Cityblock,' 'Cosine,' 'Correlation,' and 'Euclidean' distances with $K \in \{1, 2, 3, \ldots, 11\}$. For the K-means clustering algorithm, the cluster size $k \in \{4, 8, 16, \ldots, 512\}$ used for feature extraction. When increasing the histogram bin frequency value through Eq. (4), different number of minimum distance is also considered. In the first setting, the minimum distance is used. In the second one, the minimum three distances are adopted. Figure 2 provides an instance of BoW histogram construction for the five classes considered. The fact that the suggested BoW histograms are different for each of the five classes under consideration is evident from the figure. This characteristic can be treated as a notable feature for activity recognition. Hence the BoW histogram-bin frequencies are used as central features for activity recognition in this paper. To evaluate the performance of the proposed method, four evaluation metrics are used and their definitions are as follows:

$$\text{Accuracy} = \frac{T_p + T_n}{T_p + F_p + F_n + T_n} \tag{5}$$

$$\text{Precision}(P) = \frac{T_p}{T_p + F_p} \tag{6}$$

$$\text{Recall}(R) = \frac{T_p}{T_p + F_n} \tag{7}$$

$$F_1 \text{ Score} = \frac{2 \times (P \times R)}{P + R}. \tag{8}$$

Let, a signal belonging to a class 'A' is misclassified as belonging to some other class, which is defined as a false positive recognition (F_p) for class 'A'. On the other hand, a signal belonging to a different class may be misclassified as the class 'A', creating a false negative (F_n) recognition for class 'A'. If a signal is accurately predicted for class 'A', the recognition is defined as a true positive (T_p) for that class.

Also, that prediction is defined as true negative (T_n) for all other classes except class 'A'.

Besides, in order to conduct robust performance evaluation of the proposed method, k-fold cross-validation [37] has been applied. In this paper, all the results are obtained using a tenfold cross-validation scheme.

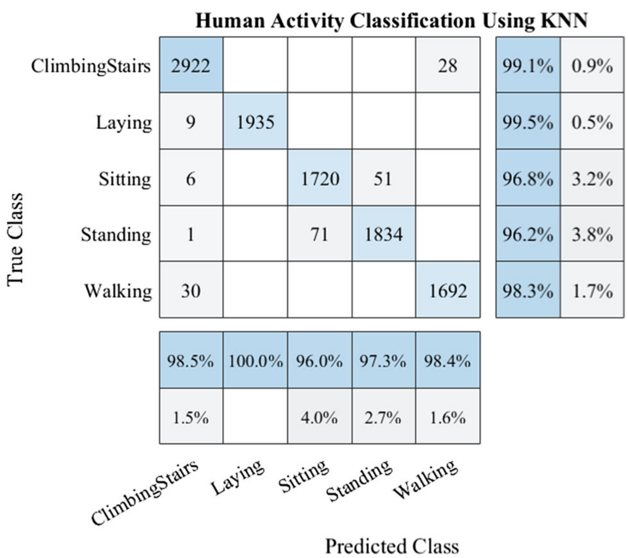

Fig. 4 The confusion matrix obtained in the case of KNN with the minimum three distances for feature extraction

Table 3 Performance comparison between the proposed method and other BoW-based models on UCI-HAR dataset

Reference	Method	Accuracy (%)
Zhang et al. [24]	Traditional BoW + KNN	94.54
Peng et al. [38]	BoW+LDA	94.45
Montero et al. [4]	Bag-of-SFA + Naive Bayes	96.85
	Bag-of-SFA + KNN	97.56
Braganca et al. [16]	Bg-of-Pattern + KNN	93.00
Proposed model	BoW + SVM	96.41
	BoW + ANN	95.78
	BoW + KNN	98.10

4.3 Results and discussion

After tuning classifier parameters, the best performance in terms of the used evaluation metrics achieved by each classifier is reported in Table 2, along with the optimal classifier settings. The overall good performance indicates the proposed feature representation can work with different classifiers very well. For comparison purpose, the results based on different combinations of classifiers and proposed feature extraction with varied cluster size for K-Means are reported in Fig. 3. As noticed, when the cluster size for K-Means is large enough, all the combinations of classifiers and proposed feature extraction could achieve reasonable good performance. When the number of minimum distance for calculating the histogram bin frequency value is fixed, KNN generates the best result compared with SVM and ANN. For the same type of classifier, it seems increasing the number of minimum distance for histogram bin frequency calculation could result in better performance. One possible explanation for this is it may share similar close distances from multiple bin centers for a sample activity signal. Among all the combinations, KNN with the minimum three distances for feature extraction leads to the best performance. The confusion matrix obtained in this case is shown in Fig. 4. It shows that laying activity is perfectly recognized by the model but there is some misclassification between walking and climbing stairs as well as between sitting and standing. Additionally, the proposed method is compared with other BoW-based methods and the results are shown in Table 3. We have implemented the algorithms presented in [24] and [38] and evaluated their performance on UCI-HAR dataset. Table 3 illustrates that our proposed method could achieve 3.56%, 3.65%, 0.54%, and 5.10% improvement compared with other BoW-based state-of-the-art models. Finally, we compared the performance of our proposed method with other state-of-the-art methods on UCI-HAR dataset, which is shown in Table 4. In consideration of the state-of-the-art especially the deep learning models, our proposed BoW-

Table 4 Performance comparison between the proposed method and other state-of-the-art models on UCI-HAR dataset

Reference	Method	Accuracy (%)	Classification time per sample
Ignatov et al. [18]	CNN	97.63	33.286–35.714 ms
Teng et al. [33]	CNN + predsim	96.98	24.462–26.055 ms
Tufek et al. [29]	3 layer LSTM	97.40	331.27–338.03 ms
Jain et al. [5]	FLF + SVM	97.12	–
Rasel et al. [28]	EPS + LDA + MCSVM	98.67	–
Xia et al. [39]	LSTM-CNN	95.78	340.62–349.80 ms
Proposed model	BoW + SVM	96.41	0.0500–0.1235 ms
	BoW + ANN	95.78	21.538–57.253 ms
	BoW + KNN	98.10	0.0036–0.0041 ms

based feature extraction scheme achieves similar or better accuracy with significantly reduced computational cost.

5 Conclusion

This paper proposes an enhanced BoW-based feature extraction algorithm for human activity recognition using sensor data from smartphones. Local features are first extracted from accelerometer and gyroscope data using DWT. Then, K-means clustering is performed on the local features for BoW vocabulary construction. To extract the discriminative BoW-based histogram feature, the bin centers corresponding to the minimum n distances from the input signal's local features are selected. Moreover, the bin frequency is increased by a trigonometry constraint cosine value of the corresponding distances. The proposed feature is evaluated using three well-known classification methods on the UCI-HAR dataset. The consistently good performance indicates the effectiveness of the proposed method. Comparative experimental results also show its superior performance than the other state-of-the-art methods in terms of cost and accuracy. In the future, we will also explore efficient deep learning models for human activity recognition.

References

1. Nweke, H.F., Teh, Y.W., Al-Garadi, M.A., Alo, U.R.: Deep learning algorithms for human activity recognition using mobile and wearable sensor networks: state of the art and research challenges. Expert Syst. Appl. **105**, 233–261 (2018)
2. Yuan, G., Wang, Z., Meng, F., Yan, Q., Xia, S.: An overview of human activity recognition based on smartphone. Sens. Rev. **39**(2), 288–306 (2019)
3. Cook, D.J., Krishnan, N.C.: Activity Learning: Discovering, Recognizing, and Predicting Human Behavior from Sensor Data. Wiley, London (2015)
4. Montero Quispe, K.G., Sousa Lima, W., Macedo Batista, D., Souto, E.: MBOSS: A symbolic representation of human activity recognition using mobile sensors. Sensors **18**(12), 4354 (2018)
5. Jain, A., Kanhangad, V.: Human activity classification in smartphones using accelerometer and gyroscope sensors. IEEE Sens. J. **18**(3), 1169–1177 (2017)
6. Chen, Z., Zhu, Q., Soh, Y.C., Zhang, L.: Robust human activity recognition using smartphone sensors via CT-PCA and online SVM. IEEE Trans. Ind. Inform. **13**(6), 3070–3080 (2017)
7. Mubashir, M., Shao, L., Seed, L.: A survey on fall detection: principles and approaches. Neurocomputing **100**, 144–152 (2013)
8. Ehatisham-ul Haq, M., Azam, M.A., Loo, J., Shuang, K., Islam, S., Naeem, U., Amin, Y.: Authentication of smartphone users based on activity recognition and mobile sensing. Sensors **17**(9), 2043 (2017)
9. Akhavian, R., Behzadan, A.H.: Smartphone-based construction workers activity recognition and classification. Autom. Constr. **71**, 198–209 (2016)
10. Yang, X., Tian, Y.: Super normal vector for human activity recognition with depth cameras. IEEE Trans. Pattern Anal. Mach. Intell. **39**(5), 1028–1039 (2016)
11. Franco, A., Magnani, A., Maio, D.: A multimodal approach for human activity recognition based on skeleton and RGB data. Pattern Recogn. Lett. **131**, 293–299 (2020)
12. Wang, A., Chen, G., Yang, J., Zhao, S., Chang, C.Y.: A comparative study on human activity recognition using inertial sensors in a smartphone. IEEE Sens. J. **16**(11), 4566–4578 (2016)
13. Margarito, J., Helaoui, R., Bianchi, A.M., Sartor, F., Bonomi, A.G.: User-independent recognition of sports activities from a single wrist-worn accelerometer: a template-matching-based approach. IEEE Trans. Biomed. Eng. **63**(4), 788–796 (2015)
14. Anguita, D., Ghio, A., Oneto, L., Parra, X., Reyes-Ortiz, J.L.: A public domain dataset for human activity recognition using smartphones. In: European Symposium on Artificial Neural Networks, Computational Intelligence and Machine Learning, ESANN (2013)
15. Gu, F., Khoshelham, K., Valaee, S., Shang, J., Zhang, R.: Locomotion activity recognition using stacked denoising autoencoders. IEEE Internet Things J. **5**(3), 2085–2093 (2018)
16. Bragancca, H., Colonna, J.G., Lima, W.S., Souto, E.: A smartphone lightweight method for human activity recognition based on information theory. Sensors **20**(7), 1856 (2020)
17. Figo, D., Diniz, P.C., Ferreira, D.R., Cardoso, J.M.: Preprocessing techniques for context recognition from accelerometer data. Pers. Ubiquit. Comput. **14**(7), 645–662 (2010)
18. Ignatov, A.: Real-time human activity recognition from accelerometer data using convolutional neural networks. Appl. Soft Comput. **62**, 915–922 (2018)
19. Ozcan, T., Basturk, A.: Human action recognition with deep learning and structural optimization using a hybrid heuristic algorithm. Cluster Computing, pp. 1–14 (2020)
20. Chen, K., Yao, L., Zhang, D., Wang, X., Chang, X., Nie, F.: A semisupervised recurrent convolutional attention model for human activity recognition. IEEE Trans. Neural Netw. Learn. Syst. **31**(5), 1747–1756 (2019)
21. Cruciani, F., Vafeiadis, A., Nugent, C., Cleland, I., McCullagh, P., Votis, K., Giakoumis, D., Tzovaras, D., Chen, L., Hamzaoui, R.: Feature learning for human activity recognition using convolutional neural networks. CCF Trans. Pervas. Comput. Interact. **2**(1), 18–32 (2020)
22. Ronao, C.A., Cho, S.B.: Human activity recognition with smartphone sensors using deep learning neural networks. Expert Syst. Appl. **59**, 235–244 (2016)
23. Howard, A.G., Zhu, M., Chen, B., Kalenichenko, D., Wang, W., Weyand, T., Andreetto, M., Adam, H.: Mobilenets: Efficient convolutional neural networks for mobile vision applications. arXiv preprint arXiv:1704.04861 (2017)
24. Zhang, Y., Jin, R., Zhou, Z.H.: Understanding bag-of-words model: a statistical framework. Int. J. Mach. Learn. Cybern. **1**(1–4), 43–52 (2010)
25. Cao, L., Wang, Y., Zhang, B., Jin, Q., Vasilakos, A.V.: Gchar: An efficient group-based context-aware human activity recognition on smartphone. J. Parallel Distrib. Comput. **118**, 67–80 (2018)
26. Acharjee, D., Mukherjee, A., Mandal, J., Mukherjee, N.: Activity recognition system using inbuilt sensors of smart mobile phone and minimizing feature vectors. Microsyst. Technol. **22**(11), 2715–2722 (2016)
27. Ahmed, N., Rafiq, J.I., Islam, M.R.: Enhanced human activity recognition based on smartphone sensor data using hybrid feature selection model. Sensors **20**(1), 317 (2020)
28. Ahmed Bhuiyan, R., Ahmed, N., Amiruzzaman, M., Islam, M.R.: A robust feature extraction model for human activity characterization using 3-axis accelerometer and gyroscope data. Sensors **20**(23), 6990 (2020)
29. Tufek, N., Yalcin, M., Altintas, M., Kalaoglu, F., Li, Y., Bahadir, S.K.: Human action recognition using deep learning methods on limited sensory data. IEEE Sens. J. **20**(6), 3101–3112 (2019)

30. Nematallah, H., Rajan, S.C., Cret, A.: Logistic model tree for human activity recognition using smartphone-based inertial sensors. Sensors (2019)
31. Irvine, N., Nugent, C., Zhang, S., Wang, H., Ng, W.W.: Neural network ensembles for sensor-based human activity recognition within smart environments. Sensors **20**(1), 216 (2020)
32. Wang, Z., Jiang, M., Hu, Y., Li, H.: An incremental learning method based on probabilistic neural networks and adjustable fuzzy clustering for human activity recognition by using wearable sensors. IEEE Trans. Inf Technol. Biomed. **16**(4), 691–699 (2012)
33. Teng, Q., Wang, K., Zhang, L., He, J.: The layer-wise training convolutional neural networks using local loss for sensor based human activity recognition. IEEE Sensors J. (2020)
34. Xiao, F., Pei, L., Chu, L., Zou, D., Yu, W., Zhu, Y., Li, T.: A deep learning method for complex human activity recognition using virtual wearable sensors. arXiv preprint arXiv:2003.01874 (2020)
35. Qin, Z., Zhang, Y., Meng, S., Qin, Z., Choo, K.K.R.: Imaging and fusing time series for wearable sensor-based human activity recognition. Inform. Fusion **53**, 80–87 (2020)
36. Kwon, Y., Kang, K., Bae, C.: Unsupervised learning for human activity recognition using smartphone sensors. Expert Syst. Appl. **41**(14), 6067–6074 (2014)
37. Wong, T.T., Yeh, P.Y.: Reliable accuracy estimates from k-fold cross validation. IEEE Trans. Knowl. Data Eng. (2019)
38. Peng, L., Chen, L., Wu, X., Guo, H., Chen, G.: Hierarchical complex activity representation and recognition using topic model and classifier level fusion. IEEE Trans. Biomed. Eng. **64**(6), 1369–1379 (2016)
39. Xia, K., Huang, J., Wang, H.: LSTM-CNN architecture for human activity recognition. IEEE Access **8**, 56855–56866 (2020)