Multi-domain Feature Extraction for Human Activity Recognition Using Wearable Sensors

Aiguo Wang
School of Electronic Information Engineering
Foshan University
Foshan, China
wangaiguo2546@163.com

Yue Meng
School of Computer Science and Technology
Donghua University
Shanghai, China
meng_yue97@163.com

Jinjun Liu, Shenghui Zhao, Guilin Chen*
School of Computer and Information Engineering
Chuzhou University
Chuzhou, China
{jinjunl, zsh, glchen}@chzu.edu.cn

2021 International Conference on Networking and Network Applications (NaNA) | 978-1-6654-4158-2/21/$31.00 ©2021 IEEE | DOI: 10.1109/NaNA53684.2021.00051

***Abstract*—The extraction and use of features from the raw sensor data plays an extremely crucial role in determining the recognition performance of an activity recognizer. Existing studies aim to train an accurate prediction model by extracting different features, however, few of them systematically investigate the power of features from different domains when they are used separately or jointly. To this end, we conduct a comparative study on multi-domain feature extraction for human activity recognition. Specifically, we first extract features from the time-, frequency-, and wavelet-domains, and then use different combinations of the three domain features to build activity recognizers. Finally, comparative experiments are performed on two activity recognition datasets and four classification models are used to avoid selection bias. Results indicate the superiority of using time-domain or frequency-domain features over wavelet features in terms of prediction performance and also show that the simultaneous use of multi-domain features generally generalizes better across datasets and classifiers, indicating that they, to a certain extent, contain complementary feature information.**

Keywords-Activity recognition; wearable sensor; time domain; frequency domain; wavelet

I. INTRODUCTION

The objective of activity recognition is to automatically recognize human activities and gain deep understanding of human behaviour, wellness, plan, and intention with the emerging technologies such as internet of things, pervasive computing, artificial intelligence, sensor technology, and big data [1]. Particularly, activity recognition essentially helps connect the raw sensor signals with practical application services in the context of ambient assisted living, human computer interaction, intelligent transportation, smart home, and among others [2]. Human activities, however, are typically associated with inherent complexity due to the nature of human behaviour. For example, there exist *inter-subject variation* and *intra-subject variation* and there are human activities that trigger similar sensor signals, making it more challenging to accurately infer human activities [3].

To handle different application scenarios and achieve enhanced recognition accuracy, researchers have proposed a wealth of feasible solutions, which we broadly group into wearable sensor-, ambient sensor-, and vision-based methods according to the used sensing units [4]. Compared with ambient sensor- and vision-based methods, wearable sensor-based methods generally have low cost, better portability, and high adherence. Besides, they tend to have a broader range of applications along with the development of various sensing devices in size and the processing power. Particularly, we can embed the sensing units such as WIFI, Bluetooth, accelerometer, camera, and gyroscope into a smartphone to better support context-aware applications [5]. According to the used models in training an activity recognizer, we can divide existing activity recognizers into knowledge-driven model and data-driven model, where the former uses domain knowledge to formally define activity specification and the latter learns an activity recognizer from the collected sensor data and generally better handles new cases.

Based on the activity recognition chain of data-driven models, the extraction and use of features from original sensor signals plays a crucial role in determining the prediction accuracy of an activity recognizer from the perspective of machine learning [1]. Even if deep learning models have the capacity to learn abstract features from raw data automatically, they generally require expensively computational resources and often suffer from limited interpretability [6]. Therefore, it is still valuable to extract hand-crafted features. In addition, studies show that hand-crafted features contain complementary information to the learned features, which helps to enhance classification performance. Accordingly, researchers have used various features to obtain the important characteristics of the collected sensor data for better discriminating different activities [7, 8], among which time-domain features, frequency-domain features, and their combinations are the most commonly used. For example, Bulling et al. extracted the mean and variance of signals and features based on fast Fourier transform (FFT) [1]. There are also studies that use wavelet transform to obtain the temporal and spectral information of raw signals [9]. For example, Preece et al. extracted wavelet features from sensor data and compared them with time-domain and frequency-domain features [9]. Although researchers have done considerable work in extracting features, few studies, to the best of our knowledge, systematically investigate the power of different domain features and their different combinations in training a robust activity recognition model.

978-1-6654-4158-2/21/$31.00 ©2021 IEEE
DOI 10.1109/NaNA53684.2021.00051

To this end, we herein perform a comparative study on the power of different domain features when they are used separately or simultaneously. The main contributions of this study include: (1) An activity recognition model based on multi-domain feature extraction is presented. Besides the use of single domain features, we also evaluate their different combinations, such as the combination of time- and wavelet-domains, and the combination of frequency- and wavelet-domains. This would potentially direct users to the extraction and use of informative features. (2) Extensive comparative experiments are conducted on two datasets. Particularly, we use four widely used classification models with different metrics to avoid the classifier selection bias. Results show that the use of multi-domain features generally obtains enhanced prediction performance across datasets and classifiers.

The reminder of this study is organized as follows. We present the activity recognition model and list the extracted features from three different domains in Section II. Section III gives experimental setup, results, and analysis, followed by the conclusion section.

II. Activity Recognition with Multi-domain Feature Extraction

A. *Activity Recognition Model*

Activity recognition chain (ARC) is mainly composed of the collection and segmentation of sensor signals, the extraction of features, the optimization of an activity recognition system, and performing activity recognition. The task of segmentation is to divide the collected time-series sensor readings into segments using sliding window techniques (e.g., pre-segment, time-based, and event-based sliding window). We then get a feature vector by extracting various features from each segment. Afterwards, we train an activity recognizer with these feature vectors and use it to predict human activities. Figure 1 presents the framework of ARC and we can integrate different components into it. Obviously, feature extraction plays an extremely crucial role in determining the performance of an activity recognizer.

From the perspective of signal representations, there are mainly three types of features for sensor-based activity recognition: time-, frequency-, and wavelet-domain features (providing temporal and spectral information). Since different domain features provide different data views, we can exploit their different combinations. That is, we can simultaneously use all of them or only use two of them (e.g., using time-domain and wavelet-domain features, using time-domain and frequency-domain features, and using frequency-domain and wavelet-domain features), where the use of time-domain and frequency-domain features is widely used in previous studies. Afterwards, we concatenate the chosen views to train an activity recognizer. It is noteworthy that other types of features such as deep learning learned features and features obtained from variational mode decomposition can also be flexibly integrated into Figure 1.

B. *Feature Extraction*

As previous studies have done, we take as time-series signals the raw sensor readings and then extract the time-, frequency-, and wavelet-domain features from the segments. Specifically, for a sensor with m axes $\{AX_1, AX_2, \ldots, AX_m\}$, we take each axis as a data source. We can extract various features from each axis and then concatenate the features of different axes to obtain the feature vector.

For time-domain features, we calculate the *standard deviation*, *mean*, *minimum*, *maximum*, *median*, *difference between the maximum and minimum*, *median absolute deviation*, *zero crossing rate*, 0.25 *quantile*, and 0.75 *quantile*. Besides, we calculate the Pearson correlation coefficient (*PCC*) between two of its axes (if $m \geq 2$). For example, for a 3-axis accelerometer, we get the *PCCs* between AX_1 and AX_2, AX_1 and AX_3, and AX_2 and AX_3.

For frequency-domain features, for each axis, we first use FFT to transform the signals into frequency domain. Afterwards, we empirically extract the *direct component*, the first five *peaks*, the *positions* of the five peaks, and *energy* that equals the sum of the squared FFT coefficients (without the direct component). Besides, according to the distribution of the FFT coefficients, we extract four *amplitude features* (including *standard deviation*, *mean*, *skewness*, and *kurtosis*) and four *shape features* (including *standard deviation*, *mean*, *skewness*, and *kurtosis*).

For wavelet features, we apply the Daubechies wavelet packet of order five (DB5) on the original time-domain signal and decompose it into five levels. Specifically, the original signal is first decomposed into a detail signal (cD1) and a coarse approximation signal. Then, the approximation signal is further decomposed into a second approximation

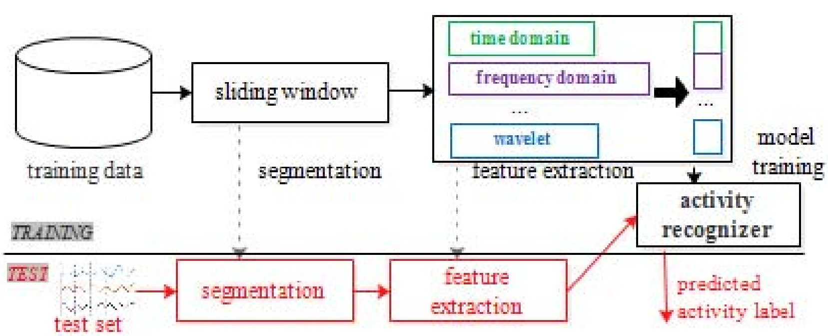


Figure 1. Activity recognition model based on multi-domain feature extraction.

signal and a detail signal (cD2). Repeat the above procedure until the defined number of decomposition level is reached. After five level decomposition, we obtain the detail signals (cD1, cD2, cD3, cD4, and cD5) at levels 1-5 and the approximation signal (cA). According to our preliminary comparative results, we take as features the sum of absolute values (i.e., $||cD1||_1$, $||cD2||_1$, $||cD3||_1$, $||cD4||_1$, $||cD5||_1$, and $||cA||_1$).

III. Experimental Setting and Results

A. Experimental Setting

We conducted comparative experiments on two public datasets to evaluate the power of time-domain, frequency-domain, wavelet features and their different combinations in recognizing human activities. The first dataset WSIDM aims to recognize six human activities (i.e., *standing*, *jogging*, *walking*, *sitting*, *upstairs*, and *downstairs*,) [7]. The sensor data was collected by a phone-based three-axis accelerometer with a 20Hz sampling rate from 29 users when they performed the activities. We segment the sensor data with a ten-second sliding window, where there is non-overlapping between two consecutive windows. We then extract time-domain, frequency-domain as well as wavelet features from each segment to represent raw sensor data. The task of the second dataset SKODA is to infer the ten manipulative gestures by using the data collected from a car maintenance assembly-line worker [8]. The sensor data sampled at 96Hz were collected by 3-axis accelerometers worn the left and right lower and upper arm of a worker. We here use one accelerometer for experiments. We first divide the sensor data with a 2s half-overlapping sliding window and then extract different domain features.

After obtaining the features, we use them separately or jointly to train activity recognizers. As for the classification models, four different classifiers, including naïve Bayes (NB), decision tree (DT), AdaBoost, and random forest (RF), are used to train the activity recognizers for comparison purposes [10, 11, 12, 13]. These models are also commonly used in previous studies. Specifically, we take decision tree as the weak learner of AdaBoost. Besides, we take as the performance metrics precision (*Prec*), recall (*Rec*), *F-measure*, and accuracy (*Acc*), where *F-measure* accounts for the imbalanced classification problems.

Given $L = \{L_1, L_2, \ldots, L_{|L|}\}$ to denote a label set with $|L|$ classes,

$$F-measure = \frac{2*precision*recall}{precision+recall}, \tag{1}$$

$$precision = \frac{1}{|L|}\sum_{i=1}^{|L|}\frac{Num_i}{NP_i}, \tag{2}$$

where Num_i is the number of samples from class L_i that are correctly classified, and NP_i is the number of samples predicted with class L_i.

$$recall = \frac{1}{|L|}\sum_{i=1}^{|L|}\frac{Num_i}{NT_i}, \tag{3}$$

where NT_i is the number of samples from class L_i.

B. Experimental Results

Tables I and II present the experimental results of the two datasets, respectively. The first column indicates the used features. Specifically, the seven entries *Time*, *Freq*, *Wav*, *Time-Freq*, *Time-Wav*, *Freq-Wav*, and *Time-Freq-Wav* correspond to the features that are from time-domain, frequency-domain, wavelet, fusion of time-domain and frequency-domain features, fusion of time-domain and wavelet features, fusion of frequency-domain and wavelet features, and combination of time-domain, frequency-domain and wavelet features. The results are organized by the used classifier, and for each group, the best accuracy is shown in bold and the second best one is shown in underlined. Besides, Figures 2 and 3 present the *F-measure* scores of the different types of features on WSIDM and SKODA, respectively.

TABLE I. Experimental results of different feature sets on WSIDM

Features	NB			DT			AdaBoost			RF		
	Acc	*Prec*	*Rec*	*Acc*	*Prec*	*Rec*	*Acc*	*Prec*	*Rec*	*Acc*	*Prec*	*Rec*
Time	76.50	73.60	73.05	93.42	90.85	91.80	98.00	96.61	98.05	98.10	96.87	98.10
Freq	84.95	82.62	81.03	94.33	91.88	92.24	97.19	95.62	96.30	97.31	95.88	96.33
Wav	81.21	78.32	78.23	89.77	85.98	86.59	94.38	90.30	94.82	94.43	90.31	94.63
Time-Freq	84.65	82.71	80.73	<u>94.49</u>	92.38	92.30	<u>98.21</u>	97.28	97.76	<u>98.29</u>	97.37	97.78
Time-Wav	78.74	78.24	77.13	93.63	91.27	91.73	97.97	96.64	98.03	98.12	96.81	98.17
Freq-Wav	**85.78**	84.26	82.24	94.28	91.92	91.91	97.43	96.14	96.72	97.34	95.96	96.53
Time-Freq-Wav	<u>85.35</u>	83.93	81.77	**94.60**	92.48	92.59	**98.25**	97.34	97.92	**98.36**	97.45	97.85

TABLE II. Experimental results of different feature sets on SKODA

Features	NB			DT			AdaBoost			RF		
	Acc	*Prec*	*Rec*	*Acc*	*Prec*	*Rec*	*Acc*	*Prec*	*Rec*	*Acc*	*Prec*	*Rec*
Time	**76.74**	78.56	79.71	87.56	86.85	86.81	95.75	95.12	95.09	96.17	95.83	95.78
Freq	63.39	66.56	69.40	84.30	84.17	84.35	93.77	93.37	93.51	93.63	93.35	93.35
Wav	60.75	63.50	62.13	77.27	76.43	76.77	88.48	87.28	87.44	87.78	86.25	86.48
Time-Freq	75.20	77.28	79.35	<u>88.44</u>	88.20	88.12	96.23	95.66	95.72	96.41	95.81	95.81
Time-Wav	76.06	77.75	78.30	87.95	87.74	87.86	<u>96.78</u>	96.07	96.03	<u>96.76</u>	95.98	95.99
Freq-Wav	70.60	73.04	73.75	85.26	84.35	84.68	95.60	94.82	94.86	95.57	94.63	94.75
Time-Freq-Wav	<u>76.41</u>	78.23	79.46	**88.79**	87.96	88.25	**96.89**	96.24	96.29	**96.96**	96.10	96.12

From Tables I and II and Figures 2 and 3, we see that the use of time-domain features and frequency-domain features outperforms the use of wavelet features except for the case of using NB on WSIDM. For example, the accuracy of using wavelet features on WSIDM for RF is 94.43% compared with the 98.10% accuracy of time-domain features and 97.31% accuracy of frequency-domain features. For SKODA, the obtained accuracies are 96.17%, 93.63%, and 87.78% for RF, respectively. Second, we observe that the use of time-domain features and the use of frequency-domain features obtain mixed results. For example, on WSIDM, the use of frequency-domain features gets higher accuracy than time-domain features with NB, while the use of time-domain features performs better with DT. Third, for the combination of two of the three different domain features, the joint use of time-domain and frequency-domain features generally has better generation ability across datasets than the combination of time-domain and wavelet features and the fusion of frequency-domain and wavelet features. Besides, we observe that the use of two different domain features obtains classification accuracy comparable to or higher than the use of single domain features. Fourth, we observe that the use of three different domain features generally obtains the best accuracy, except the case of NB (where its accuracy is comparable to the best result). For example, for RF, the use of time-domain, frequency-domain, and wavelet features gets 98.36% accuracy on WSDIM and 96.96% accuracy on SKODA. These results indicate that different domain features could complement each other and help to improve the generalization ability of an activity recognizer. Last, we observe that RF generally obtains better performance than NB, DT, and AdaBoost. This is mainly because RF, an ensemble of trees, randomly selects a subset of both samples and features to generate diverse base classification models.

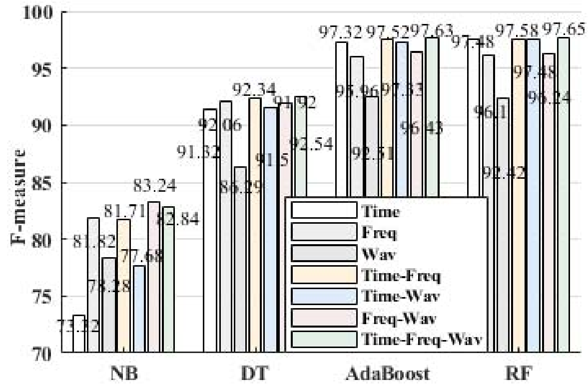


Figure 2. F-measure comparison on WSIDM.

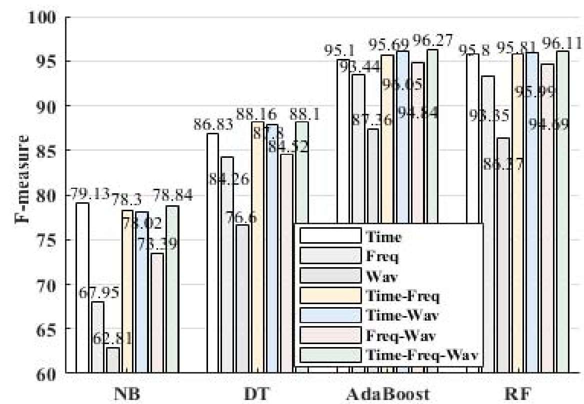
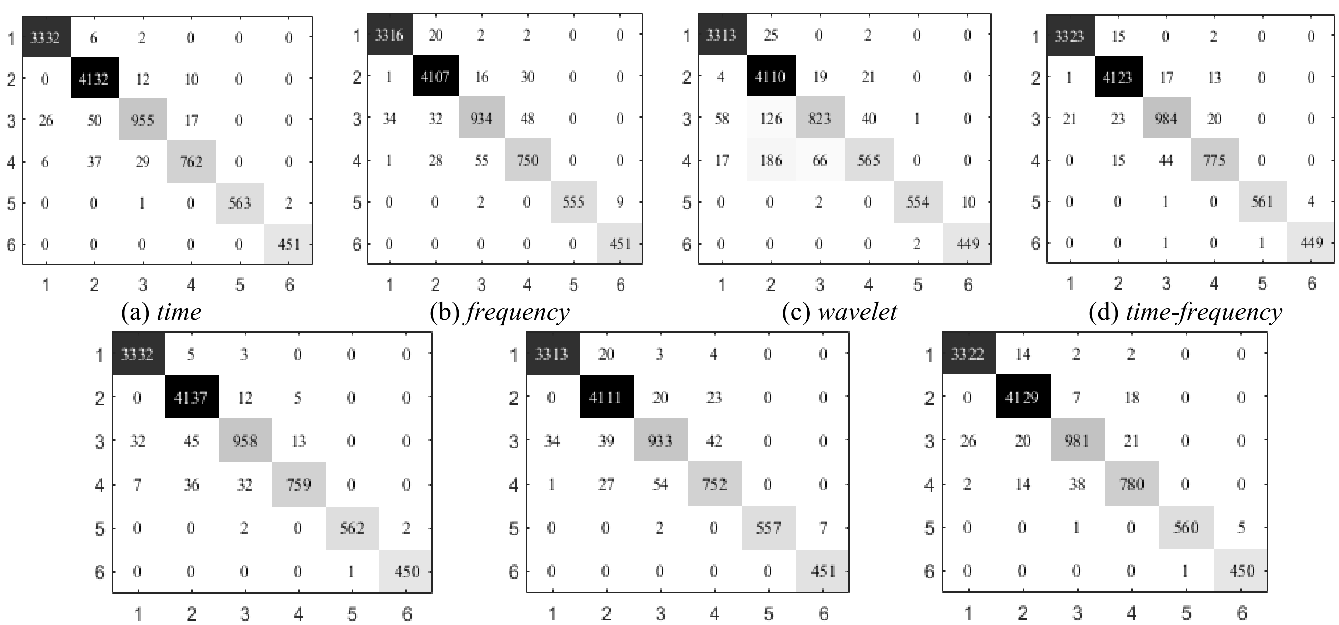


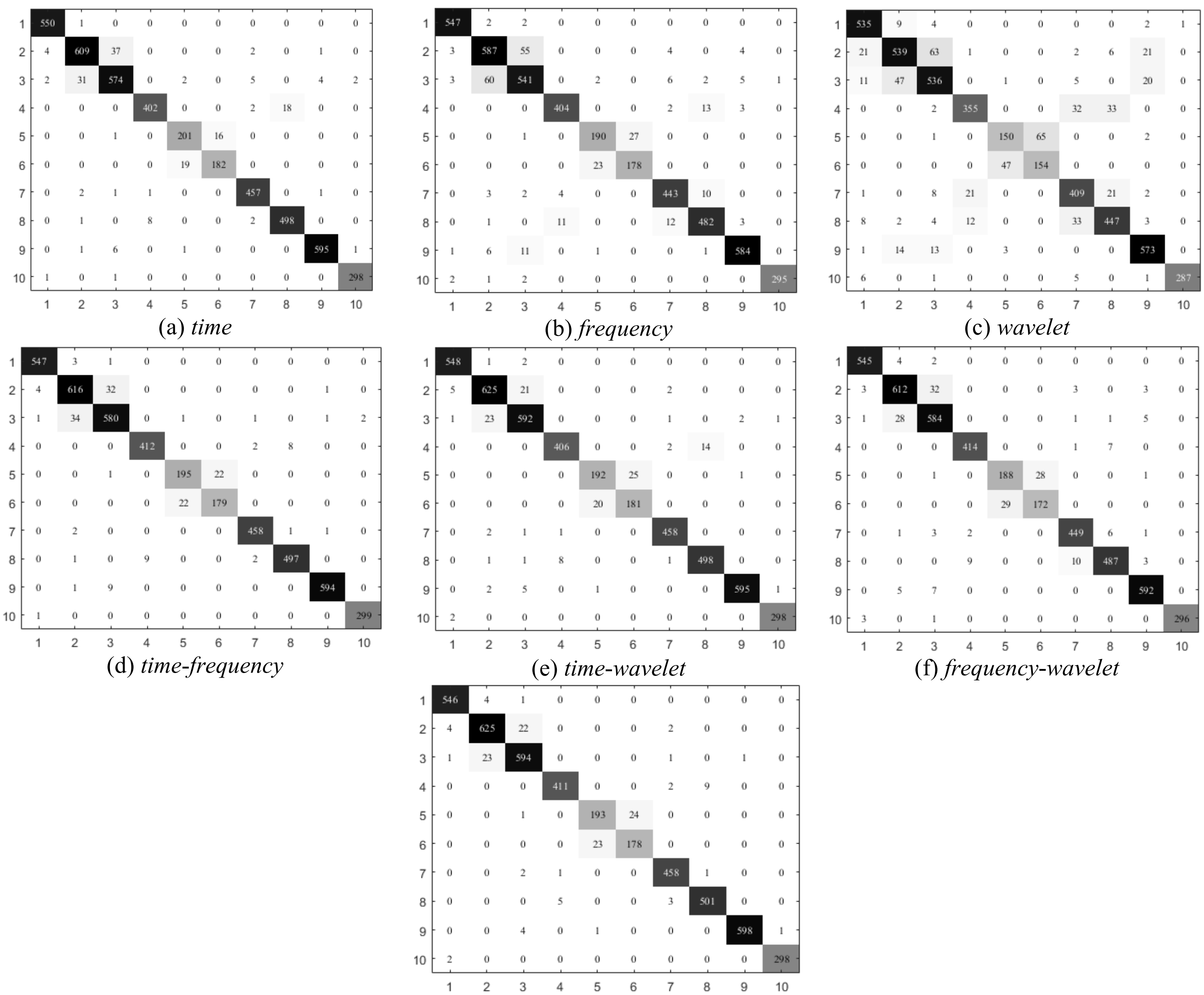
Figure 3. F-measure comparison on SKODA.



(a) *time* (b) *frequency* (c) *wavelet* (d) *time-frequency*

(e) *time-wavelet* (f) *frequency-wavelet* (g) *time-frequency-wavelet*

Figure 4. Confusion matrix on WSIDM with different types of features using RF.



(a) *time* (b) *frequency* (c) *wavelet*

(d) *time-frequency* (e) *time-wavelet* (f) *frequency-wavelet*

(g) *time-frequency-wavelet*

Figure 5. Confusion matrix on SKODA with different types of features using RF.

Furthermore, to evaluate the power of an activity recognizer in discriminating different activities, we investigate the confusion matrix. Figure 4 presents the confusion matrices on WSIDM obtained by using RF. The columns (rows) indicate the predicted (actual) labels. We see that the joint use of time-, frequency-, and wavelet-domain features outperforms its competitors. We can see that the activity *downstairs* is easily confused with the activity *upstairs* and *walking*, where the simultaneous use of three domain features only misclassifies 52 *downstairs* into *upstairs* and *walking*, compared to the 66, 83, 252, 59, 68, and 81 errors of the other six cases, respectively. Figure 5 presents the confusion matrices on SKODA of RF.

IV. CONCLUSION

From the perspective of machine learning, the extraction and use of features from the raw sensor signals determines, to a large extent, the prediction performance of an activity recognizer. There are studies that propose to extract various features from the signals, however, few studies conduct a systematic study on the power of features extracted from different domains and their combinations. To this end, we preliminarily conduct a comparative study on multi-domain feature extraction for activity recognition. Specifically, we first extract time-, frequency-, and wavelet- domain features from the sensor data and then use them separately or jointly to build activity recognizers. Finally, we conducted extensive comparative experiments on two datasets using different combinations of the three domain features. Results show that for the three domain features, the prediction performance of wavelet features is often lower than the ones of time-domain features and of frequency-domain features, and that the joint use of three domain features generally gets improved results across different datasets and classifiers. This indicates that different domain features probably provide complementary activity views.

For the future work, besides time-domain, frequency-domain, and wavelet transformation features, there are other types of features (e.g., structural features, and discrete cosine transformation features) that can be extracted from the time series signals [14, 15]. It would be meaningful to conduct a comparative study. Second, the results show that the fusion of multi-domain features may not obtain the best prediction accuracy. This is possibly because there exist redundant and irrelevant features [16, 17]. Thus, using a feature selector to optimize the feature space remains another topic.

Acknowledgment

This work was supported in part by the National Natural Science Foundation of China (Nos. 61902068, 61972092), the Featured Innovation Project of the Department of Education of Guangdong Province (No. 2021KTSCX117), the Key Research and Development Project of Anhui Province (No, 201904A05020073), the Foshan Self-funded Science and Technology Innovation Plan (Nos. 1920001001001, 1920001000529), the Natural Science Research of the Anhui Higher Education Institution (Nos. KJ2019ZD44, KJ2019A0647, KJ2020A0725), and the research project of Foshan University (No. CGZ0802502).

References

[1] A. Bulling, U. Blanke, and B. Schiele, “A tutorial on human activity recognition using body-worn inertial sensors,” ACM Comput. Surv., vol. 46, no. 3, pp. 1-33, 2014.

[2] A. Avci, S. Bosch, M. Marin-Perianu, R. Marin-Perianu, and P. Havinga, “Activity recognition using inertial sensing for healthcare, wellbeing and sports applications: A survey,” Proc. Interna. Conf. Architecture of Computing Systems, 2010, pp. 1-10.

[3] A. Wang, S. Zhao, C. Zheng, H. Chen, L. Liu, and G. Chen, “HierHAR: Sensor-based data-driven hierarchical human activity recognition,” IEEE Sensors J., vol. 21, no. 3, pp. 3353-3365, 2021.

[4] O. D. Lara and M. A. Labrador, “A survey on human activity recognition using wearable sensors,” IEEE Commun. Surv.Tut., vol. 15, no. 3, pp. 1192-1209, 2012.

[5] A. Wang, G. Chen, J. Yang, S. Zhao, and C.-Y. Chang, “A comparative study on human activity recognition using inertial sensors in a smartphone,” IEEE Sensors J., vol. 16, pp. 4566-4578, 2016.

[6] H. F. Nweke, Y. W. Teh, M. A. Al-Garadi, and U. R. Alo, “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.

[7] J. Kwapisz, G. Weiss, and S. Moore, “Activity recognition using cell phone accelerometers,” ACM SigKDD Explor. Newsl., 12(2): 74-82, 2011.

[8] P. Zappi, C. Lombriser, T. Stiefmeier, E. Farella, D. Roggen, L. Benini, and G. Tröster, “Activity recognition from on-body sensors: Accuracy-power trade-off by dynamic sensor selection,” Proc. European Conference on Wireless Sensor Networks, 2008, pp. 17-33.

[9] S. J. Preece, J. Y. Goulermas, L. Kenney, and D. Howard, “A comparison of feature extraction methods for the classification of dynamic activities from accelerometer data,” IEEE Trans. Biomed. Eng., vol. 56, no. 3, pp. 871-879, 2008.

[10] A. Wang, H. Chen, C. Zheng, L. Zhao, J. Liu, and L. Wang, “Evaluation of random forest for complex human activity recognition using wearable sensors,” Proc. 2020 Interna. Conf. on Networking and Network Applications, 2020, pp. 310-315.

[11] W. Qi, H. Su, and A. Aliverti, “A smartphone-based adaptive recognition and real-time monitoring system for human activities,” IEEE Trans. Human Mach. Syst., vol. 50, no. 5, pp. 414-423, 2020.

[12] A. Wang, N. An, G. Chen, L. Li, and G. Alterovitz, “Accelerating wrapper-based feature selection with K-nearest-neighbor,” Knowl. Based Syst., 83, 81-91, 2015.

[13] L. Peng, L. Chen, M. Wu, and G. Chen, “Complex activity recognition using acceleration, vital sign, and location data,” IEEE Trans. Mob. Comput., vol. 18, no. 7, pp. 1488-1498, Aug. 2018.

[14] J. Lin, E. J. Keogh, L. Wei, and S. Lonardi, “Experiencing SAX: A novel symbolic representation of time series,” Data Min. Knowl. Discov., vol. 15, no. 2, pp. 107–144, 2007.

[15] A. Wang, S. Zhao, C. Zheng, J. Yang, G. Chen, and C.-Y. Chang, “Activities of daily living recognition with binary environment sensors using deep learning: A comparative study,” IEEE Sensors J., vol. 21, pp. 5423-5433, 2021.

[16] A. Wang, G. Chen, X. Wu, L. Liu, N. An, and C.-Y. Chang, “Towards human activity recognition: A hierarchical feature selection framework,” Sensors, vol. 18, no. 11, p. 3629, 2018.

[17] N. Ahmed, J. I. Rafiq, and M. R. Islam, “Enhanced human activity recognition based on smartphone sensor data using hybrid feature selection model,” Sensors, vol. 20, no. 1, p. 317, 2020.