Multicolumn Bidirectional Long Short-Term Memory for Mobile Devices-Based Human Activity Recognition

Dapeng Tao, Yonggang Wen, *Senior Member, IEEE*, and Richang Hong

***Abstract*—The ever-growing popularity of mobile devices equipped with accelerometers has provided the opportunity to capture the semantic aspects of human activity and improve user experiences with behavior-based recommendations. These functions depend heavily on the accuracy of human activity recognition, and thus real applications that use mobile devices-based human activity recognition systems (MARSs) need to seamlessly incorporate the information carried by newly labeled training samples. Motivated by the success of the weightlessness feature, we propose a new two-directional feature for bidirectional long short-term memory (BLSTM) for incremental learning in human activity recognition. To further improve the performance, we also present a new ensemble classifier termed multicolumn BLSTM (MBLSTM), which effectively combines different acceleration signal features to further improve activity recognition accuracy. Experiments on the naturalistic mobile devices-based human activity dataset suggest that MBLSTM is superior to other state-of-the-art MARS methods.**

***Index Terms*—Accelerometers, bidirectional long short-term memory (BLSTM), human activity recognition, mobile service.**

I. Introduction

THE widespread adoption of mobile devices equipped with accelerometers has raised the possibility of capturing the semantics of human activity. As a result, mobile device-based human activity recognition systems (MARSs) have received increasing attention over recent years in the field of Internet of Things, including social life networks [18], [19], [44], wearables [6], [8], [37], [45], health care [1], [22], [32], and so on [23], [25], [31], [46], [48].

A critical challenge facing MARS is that real-world applications need to be seamlessly updated with information carried by newly labeled training samples. Two incremental learning [45] scenarios need to be considered in practice: first, that newly acquired samples are obtained from a person that has not previously been learned by MARS and second, the newly acquired samples are obtained from a person that has already been learned by MARS. In addition, and similar to other pattern recognition tasks, feature descriptors are important for MARS. Recently, diversified extraction methods have been proposed to improve activity classification accuracy. These have one characteristic in common, namely that the performance of the feature descriptors dictates the accuracy of activity recognition. It is of note that some studies have suggested that a satisfactory recognition rate can be achieved in accelerometer-based physical activity recognition using FFT features. However, correlations between continuous sample points are lost with FFT features.

Recently, a large number of recurrent neural network (RNN) models have been proposed to solve the sequence classification problem [14], [17], [41], [50]. RNNs allow cyclical connections and have been applied to audio and video processing and online handwriting recognition. For instance, Graves *et al.* [10] proposed bidirectional long short-term memory (BLSTM), which adequately handles long-range contextual processing and new training samples via incremental learning.

In this paper, we present a new two-directional feature for BLSTM for MARS applications. To improve performance, we also present a new ensemble classifier, multicolumn BLSTM (MBLSTM). The architecture of the proposed MARS based on MBLSTM is shown in Fig. 1. Briefly, we first collect a large number of labeled samples to build an initial expandable model using offline learning, and second, users further update the initial model via incremental learning.

This paper is organized as follows: in Section II, we present related works on image annotation; in Section III, we detail the newly proposed MARS; in Section IV, we present the experimental results of the new approach applied to the naturalistic mobile devices-based human activity (NMHA) dataset; and we conclude in Section V.

II. Related Work

As in other pattern recognition applications such as image classification [24], [43], [49] and character recognition [36], two important stages need to be considered in MARS: feature extraction and classification.

Manuscript received March 21, 2016; accepted April 15, 2016. Date of publication May 3, 2016; date of current version January 10, 2017. This work was supported in part by the National Natural Science Foundation of China under Grant 61572486 and Grant 61472116, in part by the Guangdong Natural Science Funds under Grant 2014A030310252, in part by the Shenzhen Technology Project under Grant JCYJ20140901003939001, and in part by the Singapore MOE Tier-1 under Grant RG13/17.

D. Tao is with the School of Information Science and Engineering, Yunnan University, Kunming 650091, China (e-mail: dapeng.tao@gmail.com).

Y. Wen is with the School of Computer Engineering, Nanyang Technological University, Singapore 639798 (e-mail: ygwen@ntu.edu.sg).

R. Hong is with the Department of Electronic Engineering and Information System, Hefei University of Technology, Hefei 230009, China (e-mail: hongrc@hfut.edu.cn).

Digital Object Identifier 10.1109/JIOT.2016.2561962

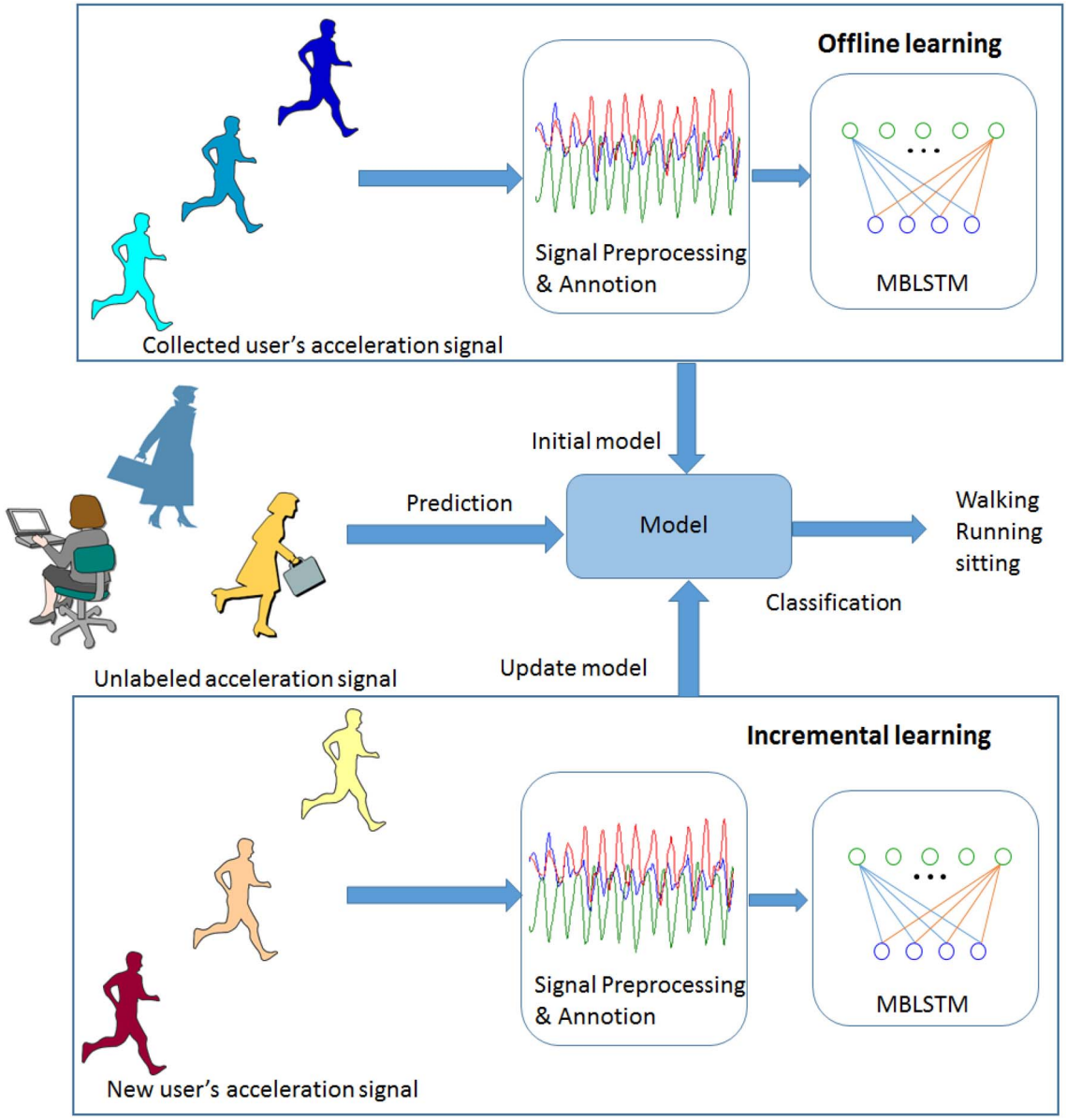


Fig. 1. MARS.

Extracting appropriate features from raw data is the most critical part of activity recognition [2], [39], [47]. Diversified feature extractors can be applied after the acceleration signal is segmented, and various feature extracting approaches have been proposed. The most commonly used feature extraction methods are based on time-domain and frequency-domain analyses.

Time-domain analyses usually consider features that can be extracted directly from raw data along a time axis, and a very large number of these have been proposed for activity recognition systems. Early activity recognition researchers introduced mean values to their studies [7], [33], in which the mean value of the acceleration data inside a window is used to represent the raw sensor data by removing random spikes and different kinds of noise. Furthermore, the computational costs and memory requirements of mean-value methods are very low, further popularizing the technique. Other simple statistical parameters such as maximum and minimum values have also be used for activity recognition (see [12], [20]). For instance, Farringdon *et al.* [7] used the difference between maximum and minimum values with other indicators to distinguish between walking and running. Variance and standard deviation have also been used as feature extractors with some success [27].

With respect to frequency-based analysis, most studies have relied on Fourier transforms to decompose signals into their different constituent frequencies, with accordingly good gains in the performance of activity recognition. Since conventional Fourier transform components are complex numbers, it is reasonable to consider their magnitudes [29]. The energy, which is the normalized squared sum of FFT components, was also used in [29] to help detect different activities. However, the energies of two distinct activities may be very similar and confuse the system. In these cases, the concept of entropy can be

applied to further differentiate energy features [16]. In addition, linear FFT frequency bands, log FFT frequency bands, and cepstral coefficients are useful frequency features [21]: the first is the real valued FFT values grouped into linear bands, the second describes the logarithmic bands that categorize real valued FFT values, and cepstral coefficients are calculated from the Fourier transform of the log FFT spectrum.

After feature extraction and compaction, various classifiers can be applied to the system to accurately recognize activities. A number of well-established classification tools have been used for this purpose, including k-nearest neighbors (k-NNs), naïve Bayes, support vector machines (SVMs), multilayer perceptron (MLP; a neural network), and k-means.

Decision trees are famous for their simplicity but strength in analyzing input variables, such as in [11] and [38]. The k-NN algorithm is another simple but powerful method to predict the class of a target based on training samples, and is implemented by computing the distances between target and training samples and finding the k nearest samples and deciding the target label using a majority vote. Although k-NN is simple in concept, the classification accuracy is still relatively good [3], [28], [42]. Naïve Bayes classifiers assume that features are conditionally independent, and the relationship between the target value and each feature is learned according to Bayes' theorem; naïve Bayes has been used in [5] and [28] for activity recognition. SVMs are more sophisticated but are widely used and the theory is well established, and SVMs have been used for activity recognition [26], [34]. Finally, MLP has advantages for activity recognition since it can learn the deep relationship between the input and output vectors and the system does not require elusive mathematical procedures [13], [30].

III. MARS Framework

In this section, we present a new MARS.

A. Two-Directional Features

Features play a critical role in supervised learning schemes. Motived by the success of the weightlessness feature [15], we have divided acceleration into its horizontal and vertical components to produce a novel two-directional feature. In addition, since the acceleration peaks might better reveal activity patterns [15], it is reasonable to cut out a section of the signal using a fixed-length window centered on a peak. Prior to peak detection, a moving average filter is used on each axis' sample points. Further details of the feature extraction steps are provided below.

In human activity categorization, one sampling window output of a three-axis accelerometer represented by a matrix $\{A_i\}_{i=1}^{N} = [A_1, A_2, \ldots, A_N]$ consists of N points, with each sample point A_i a 3-D vector

$$A_i = \left(a_{xi}, a_{yi}, a_{zi}\right). \tag{1}$$

Considering the accelerometer outputs 125, the acceleration measurement is $0g$ (where $g = 9.8\text{m/s}^2$). We define the constant vector

$$M = \left(m_{x0}, m_{y0}, m_{z0}\right) = (125, 125, 125). \tag{2}$$

Thus, the measurements of three-axis accelerometers are

$$A_i^{'} = A_i - M = \left(a_{xi} - m_{x0}, a_{yi} - m_{y0}, a_{zi} - m_{z0}\right). \tag{3}$$

The mean values of the acceleration signal are calculated by using

$$\overline{A_x^{'}} = \frac{\sum_{i=1}^{N}(a_{xi} - m_{x0})}{N} \tag{4}$$

$$\overline{A_y^{'}} = \frac{\sum_{i=1}^{N}\left(a_{yi} - m_{y0}\right)}{N} \tag{5}$$

$$\overline{A_z^{'}} = \frac{\sum_{i=1}^{N}\left(a_{zi} - m_{y0}\right)}{N}. \tag{6}$$

The acceleration is composed of gravitational acceleration and activity-based acceleration, and we assume that the mean value of the activity-based acceleration is very small over a period of time. Thus, the vertical direction is the same as gravitational direction can be calculated with

$$e_V = \frac{\left(\overline{A_x^{'}}, \overline{A_y^{'}}, \overline{A_z^{'}}\right)}{\left\|\left(\overline{A_x^{'}}, \overline{A_y^{'}}, \overline{A_z^{'}}\right)\right\|}. \tag{7}$$

The vertical component of acceleration can be calibrated using

$$V_i = A_i^{'} \cdot e_V \tag{8}$$

and the norm of the horizontal component can be calculated with

$$H_i = \left\|A_i^{'} - V_i\right\|. \tag{9}$$

Since a human activity receives a combination effect of horizontal and vertical force, Two-directional features that contains V_i and H_i can be used to characterize the relationships between activities and acceleration. The steps of two-directional features are shown in Fig. 2. In addition, the norm of the measurements of three-axis accelerometers $\|A_i^{'}\|$ is also a feature descriptor.

B. BLSTM Classifier

Graves *et al.* [10] demonstrated that BLSTM handles data with long-range interdependencies extremely well. Here, we briefly describe the method.

Considering an input sequence $\{x^t\}$, BLSTM has two scanning directions as shown in Fig. 3(a). We define a_h^t as the network input to LSTM unit h at time t, and b_h^t is the activation of the LSTM unit h at time t. Let w_{lh} be the weight of the feed-forward connection from input unit l to hidden unit h. $w_{h'h}$ is the weight of the recurrent connection from hidden unit $h^{'}$ to unit h.

The forward equations for BLSTM with L input units and H hidden summation units are

$$a_h^t = \sum_{l=1}^{L} x_l^t w_{lh} + \sum_{\substack{h^{'}=1 \\ t>0}}^{H} b_{h'}^{t-1} w_{h'h} \tag{10}$$

and

$$b_h^t = \theta_h\left(a_h^t\right) \tag{11}$$

where θ_h is the activation function of hidden unit h.

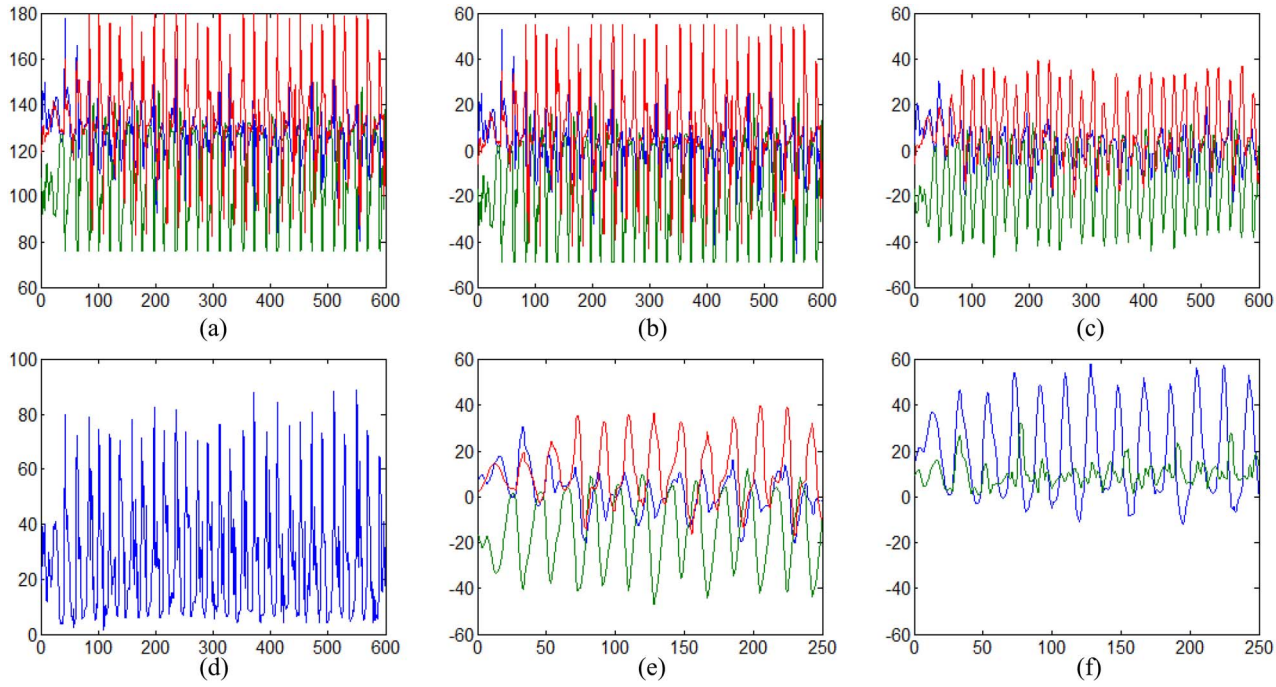


Fig. 2. Steps of two-directional features. (a) Output of three-axis accelerometer. (b) Measurements of three-axis accelerometers. (c) After moving average filter. (d) $\|A_i'\|$. (e) 250 points A_i. (f) Two-directional features.

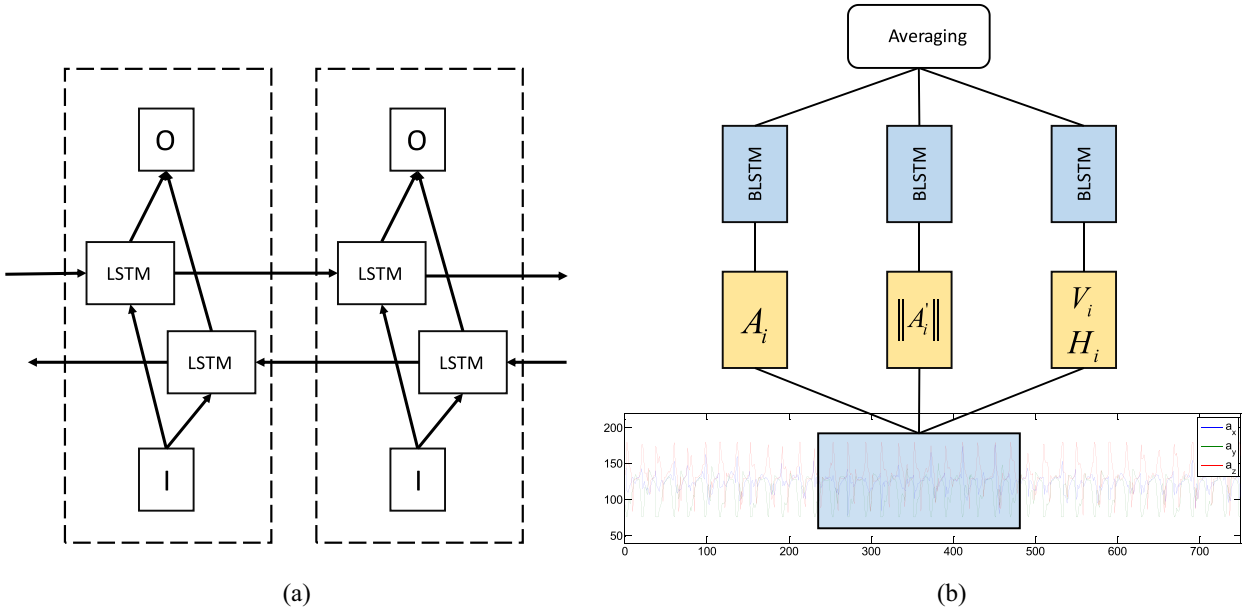


Fig. 3. (a) Scanning directions of BLSTM, where "*I*" refers to input and "*O*" refers to output. (b) Network architecture of MBLSTM.

Let O be the objective function, which can be defined using cross-entropy error. The backward equations for BLSTM with K output units and H hidden summation units are

$$\frac{\partial O}{\partial w_{hk}} = \sum_{t=1}^{T} \frac{\partial O}{\partial a_h^t} \frac{\partial a_h^t}{\partial w_{hk}} = \sum_{t=1}^{T} \frac{\partial O}{\partial a_h^t} b_h^t \tag{12}$$

where

$$\frac{\partial O}{\partial a_h^t} = \theta_h'(a_h^t) \left(\sum_{k=1}^{K} \frac{\partial O}{\partial a_k^t} w_{hk} + \sum_{h'=1t>0}^{H} \frac{\partial O}{\partial a_{h'}^{t+1}} w_{hh'} \right). \tag{13}$$

Due to space constraints, the other parts of BLSTM (the input gate, forget gate, output gate, and the memory cell) are not detailed here but are easy to implement; the interested reader is referred to [9] and [10].

In our classification problem, hierarchical design choices, i.e., the outputs of one layer used as the inputs to the next layer, are suitable for feature selection. The network architecture therefore comprises an input layer, two BLSTM layers, a feed-forward layer, a collapse layer, and an output layer.

The procedure is as follows.

1) The input sequence is partitioned into blocks of size 3. Each block is scanned into a vector as a single input of the first hidden layer, i.e., BLSTM, which has four units and scans the input blocks in two directions.
2) The activations of the first hidden layer are given as the input to the second hidden layer, i.e., the feed-forward layer, which consists of six units.
3) The third hidden layer, i.e., BLSTM, has ten units and scans the activations of the second hidden layer in two directions.
4) The forth hidden layer is a collapse layer, which sums over all inputs from the third hidden layer at each time step.
5) The output of the last map is fed into an N-way softmax, where N is the number of activity categories.

C. Architecture of MBLSTM

Inspired by multicolumn DNN [4], we combine several BLSTM classifiers to form an MBLSTM. Specifically, different feature descriptors are independently trained on the same BLSTM model version, with democratically averaged predictions then performed with them. Fig. 3(b) shows MBLSTM's network architecture, which comprises three BLSTMs. From the left to right classifier, the feature descriptor is A_i, $\|A_i^{'}\|$, and two-directional features, respectively.

IV. Experimental Results

We next conducted mobile devices-based human activity recognition experiments to demonstrate the effectiveness of the proposed MBLSTM. The NMHA dataset, which was collected from 100 subjects utilizing smart phones equipped with three-axis accelerometers, is used in our experiments. Comparative experiments were conducted using different algorithms on the same dataset to fully evaluate MBLSTM.

We measured system performance by calculating the average accuracy for each human activity category. In addition, a confusion matrix was used to illustrate where the method failed. The details of experimental setup and baseline methods are presented below.

A. Datasets

There is a lack of standardized datasets for mobile devices-based human activity recognition. Several studies have evaluated performance on the SCUT naturalistic 3-D acceleration-based activity (SCUT-NAA) dataset and the physical activity monitoring for aging people (PAMAP) dataset [40]. However, these are not suitable for our experiments because: 1) the data are not collected from smart phones and 2) there are insufficient activity samples from different people. The SCUT-NAA dataset consists of 1278 samples obtained from 44 people (34 males and 10 females), while the PAMAP dataset collected the raw acceleration data from only nine subjects. Thus, we collected 100 human behavior activity samples to create a dataset that utilized smart phones equipped with a 3-D accelerometers, which we call the NMHA dataset. We cut out a section of the acceleration signal using a 250-point window centered on a signal peak to obtain an activity sample. The distribution of human activity samples is shown in Table I, and example acceleration signals are shown in Fig. 4.

TABLE I
Statistics of Samples

Classes	Subjects	Sample size
Jumping	100	5835
Running	100	4868
Walking	100	3152
Step walking	100	1254
Walking quickly	100	3438
Down stairs	100	4650
Up stairs	100	3270

B. Experiment Results on Two-Directional Features

In order to evaluate two-directional features, we compared them with sample point A_i and the norm of the measurements of three-axis accelerometers $\|A_i^{'}\|$. In our experiments, we randomly selected all samples from $p_{tr} = 50, 60, 70$, and 80 subjects to form the training set and randomly selected all samples from $p_{ts} = 20$ subjects to form the test set. In addition, we composed our validation set with a fifth of the training set. The validation set was used to tune important parameters such as the learning rate. In particular, the learning rate was divided by 10 when 30 time points had passed without any reduction in error on the validation set with the current learning rate. In addition, for the BLSTM classifier, BLSTM was trained using OGD with a learning rate of 10^{-4} and a momentum of 0.9 during training. The validation error was assessed after each round of training, and the termination criterion for training was judged by the new lowest value for the validation error in ten iterations.

The test error rates of three features descriptors are shown in Tables II–IV. In addition, the training and validation error rates when training was stopped are shown. It can be seen that the two-directional features are a promising feature descriptor for the BLSTM scheme, because it can extract two important components for MARS, i.e., the horizontal and vertical components. In order to inspect the function of the validation set, we also depicted BLSTM activity recognition error rates during training (Figs. 5–7). It can be seen that BLSTM using two-directional features converges faster, which is extremely important for practical applications.

C. Experiment Results on MBLSTM

We next conducted experiments to evaluate the effectiveness of MBLSTM and compare it with several popular methods including SVM [51], EMR [35], k-NN [28], and BLSTM. In our experiments, we randomly selected all samples from $p_{tr} = 40$ and 80 subjects to form the training set, while the remaining subjects were used as the test set. We composed

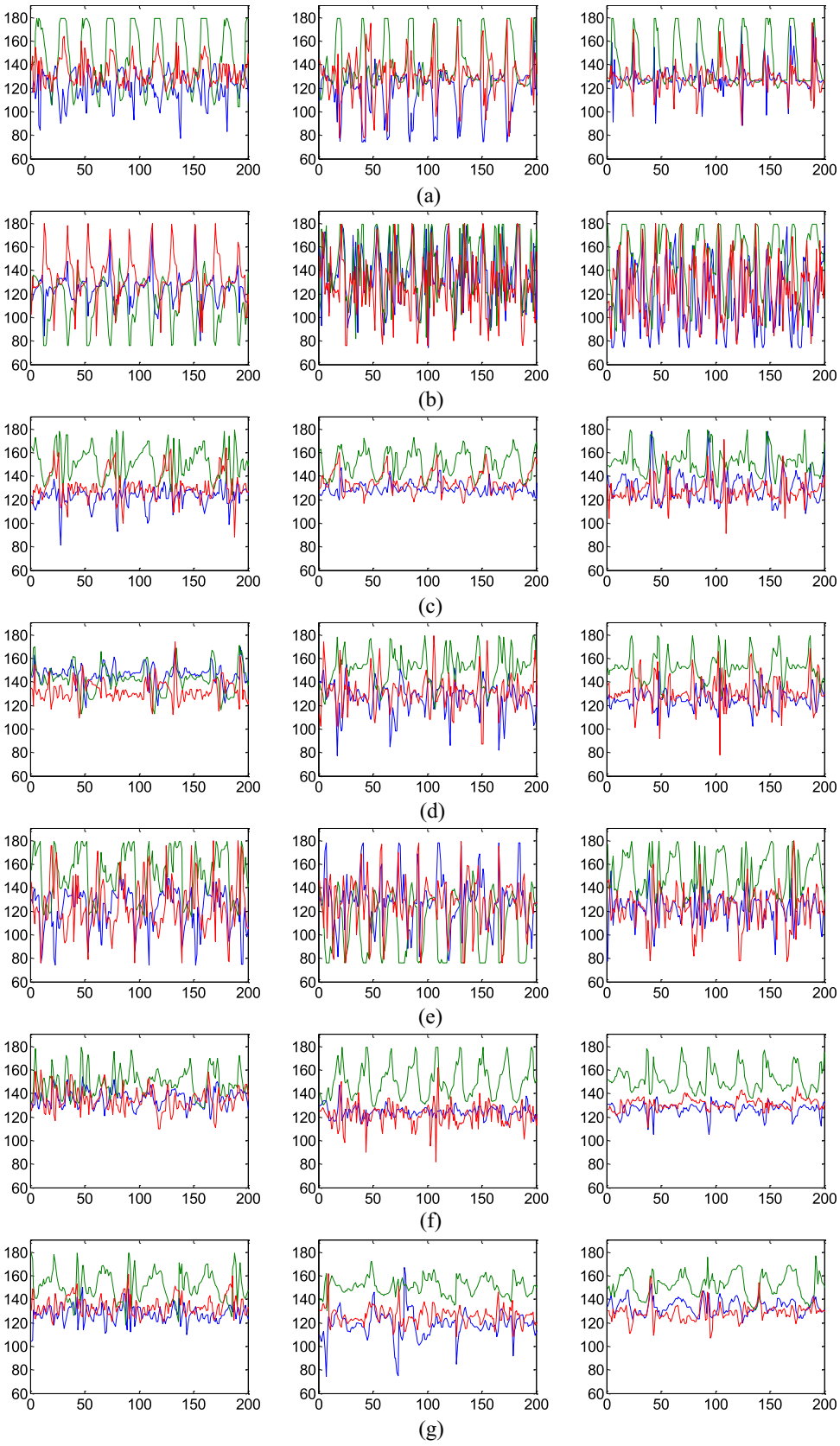


Fig. 4. NMHA dataset samples. (a) Jumping. (b) Running. (c) Walking. (d) Step walking. (e) Walking quickly. (f) Down stairs. (g) Up stairs.

our validation set with a fifth of the training set. The process was repeated five times and then averaged the average error rates for each class.

Comparisons between MBLSTM and SVM, EMR, *k*-NN, and BLSTM on the NMHA dataset are shown in Table V. For SVM, EMR, and *k*-NN classifiers, we concatenate the

TABLE II
EXPERIMENT RESULTS ON A_i FEATURE DESCRIPTOR

Classes	Training error rate (%)				Validation error rate (%)				Test error rate (%)			
	$p_{tr}=50$	$p_{tr}=60$	$p_{tr}=70$	$p_{tr}=80$	$p_{tr}=50$	$p_{tr}=60$	$p_{tr}=70$	$p_{tr}=80$	$p_{tr}=50$	$p_{tr}=60$	$p_{tr}=70$	$p_{tr}=80$
Jumping	6.2	4.5	7.2	1.1	15.4	20.1	4.6	5.3	29.5	17.7	5.9	4.8
Running	10.1	8.2	12.9	1.2	15.9	18.7	23.1	14.7	13.3	25.1	12.8	9.5
Walking	52.6	53.9	48.3	20.1	66.4	45.1	38.6	32.4	69.4	60.4	62.8	46.3
Step walking	55.6	48.0	59.4	15.2	64.9	67.7	71.6	19.8	76.2	71.7	82.3	26.0
Walking quickly	45.4	38.9	26.9	15.8	52.4	48.3	49.1	48.8	58.0	53.4	37.3	32.7
Down stairs	20.6	19.2	19.7	4.3	39.6	41.7	43.5	22.8	33.3	28.1	36.2	16.9
Up stairs	57.3	41.8	34.5	9.6	48.5	57.8	53.9	35.8	66.6	70.1	64.7	22.9
Total error	28.4	24.8	24.1	7.6	37.6	37.3	33.6	23.6	43.0	40.2	34.8	19.6

TABLE III
EXPERIMENT RESULTS ON $\|A_i\|$ FEATURE DESCRIPTOR

Classes	Training error rate (%)				Validation error rate (%)				Test error rate (%)			
	$p_{tr}=50$	$p_{tr}=60$	$p_{tr}=70$	$p_{tr}=80$	$p_{tr}=50$	$p_{tr}=60$	$p_{tr}=70$	$p_{tr}=80$	$p_{tr}=50$	$p_{tr}=60$	$p_{tr}=70$	$p_{tr}=80$
Jumping	5.2	2.3	4.5	1.4	6.8	3.1	4.2	7.6	2.6	5.3	0.5	1.7
Running	13.6	5.9	7.8	2.3	4.0	9.3	11.2	5.0	11.5	13.4	16.6	14.2
Walking	57.0	29.5	32.1	14.7	55.4	52.4	37.5	20.9	67.2	45.6	41.0	30.7
Step walking	91.5	38.2	22.3	3.7	84.4	60.9	15.8	12.4	93.2	52.8	37.4	24.2
Walking quickly	42.6	26.1	23.1	18.3	50.4	42.7	43.8	50.2	39.7	46.6	35.6	32.2
Down stairs	25.8	7.3	11.1	4.6	23.1	18.8	24.4	15.7	21.3	16.9	21.3	15.0
Up stairs	32.0	24.0	28.6	10.7	67.4	39.8	37.6	27.9	55.4	31.3	42.2	26.4
Total error	28.1	14.5	16.0	7.1	32.6	25.9	22.8	18.9	31.6	24.6	23.6	17.9

TABLE IV
EXPERIMENT RESULTS ON TWO-DIRECTIONAL FEATURES

Classes	Training error rate (%)				Validation error rate (%)				Test error rate (%)			
	$p_{tr}=50$	$p_{tr}=60$	$p_{tr}=70$	$p_{tr}=80$	$p_{tr}=50$	$p_{tr}=60$	$p_{tr}=70$	$p_{tr}=80$	$p_{tr}=50$	$p_{tr}=60$	$p_{tr}=70$	$p_{tr}=80$
Jumping	1.7	2.8	2.6	1.3	5.9	1.8	1.2	1.3	2.4	2.4	0.8	1.5
Running	4.2	3.4	3.5	1.7	1.4	4.1	6.1	5.8	6.4	4.6	9.1	2.9
Walking	23.8	11.3	12.1	16.2	29.8	8.7	17.8	21.7	41.2	38.0	34.2	35.2
Step walking	14.3	7.9	5.3	4.6	14.9	3.1	1.1	0.5	40.4	29.8	21.5	20.4
Walking quickly	23.0	12.4	13.4	14.8	31.3	40.5	45.4	37.7	29.8	38.0	33.1	23.2
Down stairs	9.7	4.7	3.1	4.6	11.7	20.7	16.5	15.7	13.5	13.7	11.7	11.5
Up stairs	17.9	11.1	7.7	7.1	21.5	34.4	26.5	25.5	30.5	30.5	27.9	16.7
Total error	11.4	6.7	6.2	6.3	15.0	16.1	16.1	14.9	18.8	18.5	16.8	13.2

two-directional features to a long feature vector, and use the long feature vector in above classifiers. And the two-directional features are used in BLSTM classifier. From Table V, we can observe that the proposed two-directional features are also suitable for other classifiers. MBLSTM and BLSTM represent promising solutions to MARS because they model the correlations between continuous sample points of the acceleration signal. In addition, combining several BLSTM columns into an MBLSTM further decreases the error rate to close to 20%.

The mean time for testing is 2.7 ms, and the test step is on the cloud. We conduct all experiments on an Intel Xeon CPU E5-2640 v2 @ 2.0 GHz computer with a 32 GB memory. The classification confusion matrix of MBLSTM for one test split of the NMHA dataset is shown in Fig. 8, with correct labels depicted on the vertical axis and predicted labels on the horizontal axis. It is convenient to inspect the pattern of classification errors, since they are nonzero values listed outside the diagonal. Unsurprisingly, confusions occur between "walking," "step walking," and "walking quickly" because there are very similar.

D. Experiment Results on Incremental Learning

The first instances of incremental learning were demonstrated in Section IV-B. Here, we design an incremental learning experiment considering a second situation.

The procedure is as follows.

1) First, we randomly selected $p_{tr} = 40$ and 80 people to establish the training set and the remaining people were used to form the test set.

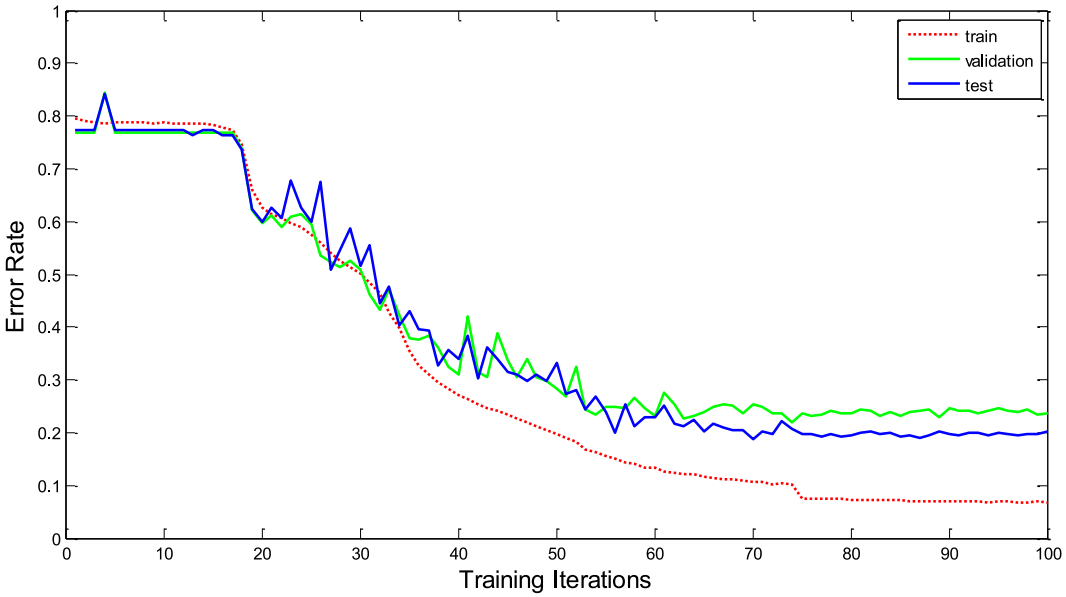


Fig. 5. BLSTM recognition error rate during training. We applied the A_i feature descriptor as the input of BLSTM.

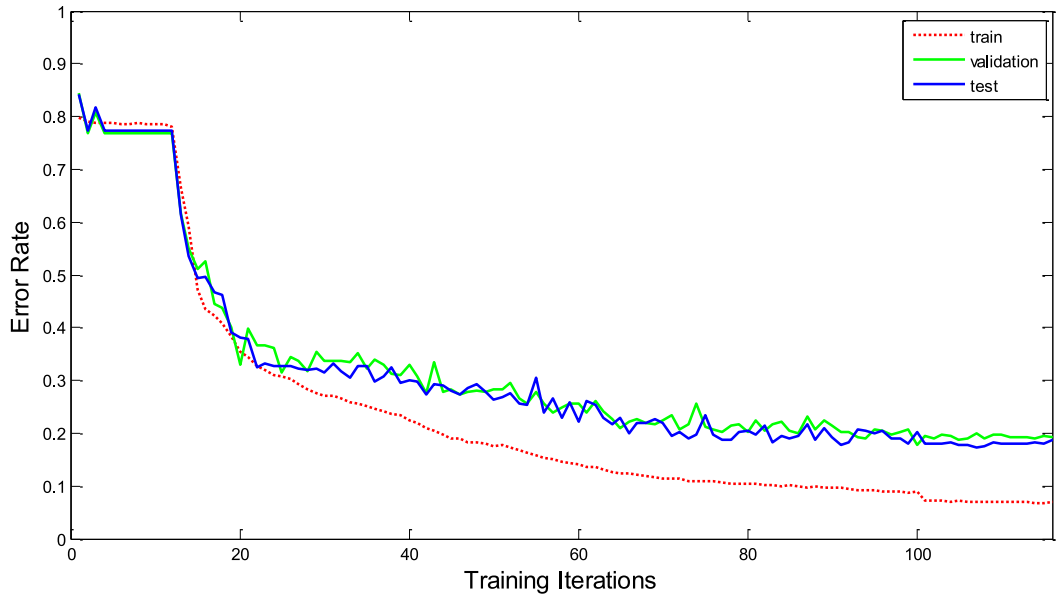


Fig. 6. BLSTM recognition error rate during training. We applied the $\|A_i'\|$ feature descriptor as the input of BLSTM.

TABLE V
AVERAGE ERROR RATES OF FIVE ALGORITHMS

Class	$p_{tr}=40$					$p_{tr}=80$				
	SVM	EMR	kNN	BLSTM	MBLSTM	SVM	EMR	kNN	BLSTM	MBLSTM
Jumping	1.7	1.7	5.2	4.1	1.3	3.8	1.2	3.7	1.5	0.0
Running	6.7	10.7	13.0	6.8	1.2	9.1	11.4	11.5	2.9	3.9
Walking	46.9	39.5	41.7	40.6	39.9	22.7	22.9	33.1	35.2	26.3
Step walking	57.0	40.2	58.9	17.7	20.0	32.1	31.2	30.9	20.4	17.7
Walking quickly	31.2	32.7	36.4	40.2	33.6	34.9	32.6	40.2	23.2	23.2
Down stairs	11.3	14.5	16.5	13.9	5.4	11.2	15.2	18.3	11.5	6.8
Up stairs	37.3	35.8	36.9	29.0	29.2	24.2	25.9	31.3	16.7	13.9
Average error rates	21.0	20.4	23.8	19.1	15.4	16.2	16.6	20.6	13.2	10.6

2) Second, we randomly selected $p_{ti} = 1, 2, 3, 4$, and 5 samples from each person in the test set and applied MBLSTM to incrementally update the model obtained by the training set.

3) The process was repeated five times and the error rates averaged.

The results of incremental learning are shown in Table VI. It can be seen that newly added samples significantly decrease

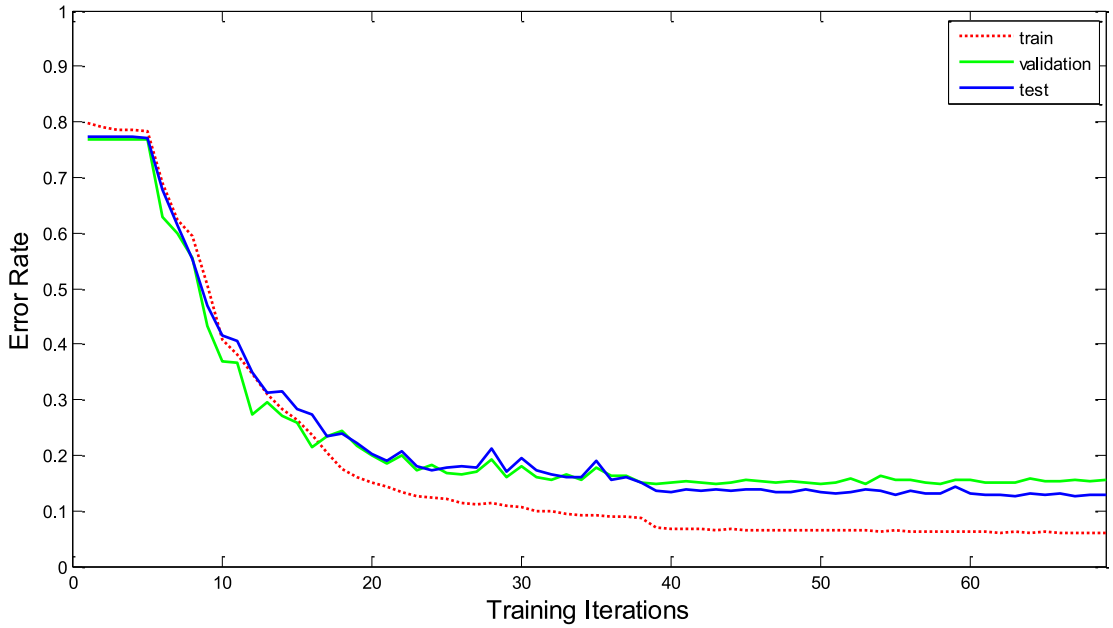


Fig. 7. BLSTM recognition error rate during training. We applied the two-directional features as the input of BLSTM.

TABLE VI
EXPERIMENT RESULTS ON INCREMENTAL LEARNING

Class	$p_{tr}=40$					$p_{tr}=80$				
	$p_{ti}=1$	$p_{ti}=2$	$p_{ti}=3$	$p_{ti}=4$	$p_{ti}=5$	$p_{ti}=1$	$p_{ti}=2$	$p_{ti}=3$	$p_{ti}=4$	$p_{ti}=5$
Jumping	0.3	0.2	0.0	0.0	0.0	0.0	0.1	0.0	0.0	0.0
Running	0.2	0.2	0.0	0.0	0.0	3.9	2.4	0.8	0.4	0.0
Walking	29.3	19.1	16.6	4.2	3.9	15.3	7.9	6.1	3.0	3.9
Step walking	15.9	12.8	11.5	11.6	10.7	12.6	11.0	7.7	10.6	11.9
Walking quickly	21.2	10.5	10.1	2.7	1.4	19.3	8.9	6.7	2.5	3.9
Down stairs	1.8	2.3	1.9	1.2	1.0	4.9	2.2	0.6	2.1	1.4
Up stairs	15.2	6.9	6.6	4.9	7.4	4.3	3.2	2.4	1.6	0.9
Average error rates	9.4	5.6	4.9	2.2	2.3	6.9	3.8	2.5	1.8	1.8

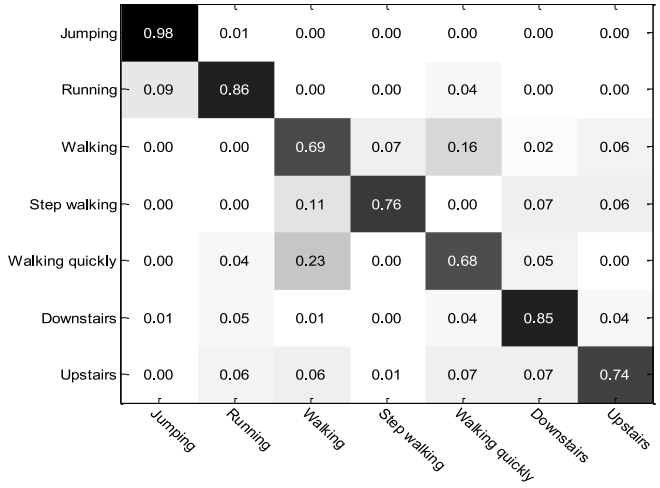


Fig. 8. MBLSTM classification confusion matrix for one test split on the NMHA dataset. Correct labels are on the vertical axis; predicted labels are on the horizontal axis.

the error rates. Thus, the user can achieve a better experience by labeling a small amount of data on mobile devices in real applications.

V. CONCLUSION

Smart phones equipped with three-axis accelerometers are becoming increasingly popular. Thus, mobile devices-based human activity recognition is becoming an important application, and has already contributed to developments in social media. Considering the complexity of the three-axis acceleration signal, here we present a new ensemble classifier: MBLSTM to build our MARS. By studying the original acceleration signal, we also present a suitable feature descriptor, the two-directional feature, as a BLSM classifier unit. Seamless integration of two-directional features and MBLSTM significantly improves activity recognition in our experiments.

REFERENCES

[1] L. Bao and S. S. Intille, "Activity recognition from user-annotated acceleration data," in *Proc. 2nd Int. Conf. Pervasive Comput.*, Vienna, Austria, 2004, pp. 1–17.

[2] O. Banos, M. Damas, H. Pomares, A. Prieto, and I. Rojas, "Daily living activity recognition based on statistical feature quality group selection," *Expert Syst. Appl.*, vol. 39, no. 9, pp. 8013–8021, 2012.

[3] M. Benocci *et al.*, "Wearable assistant for load monitoring: Recognition of on—Body load placement from gait alterations," in *Proc. Int. Conf. Pervasive Comput. Technol. Healthcare*, Munich, Germany, 2010, pp. 1–8.

[4] D. Ciresan, U. Meier, and J. Schmidhuber, "Multi-column deep neural networks for image classification," *Comput. Vis. Pattern Recognit.*, pp. 3642–3649, Feb. 2012.

[5] S. Dernbach, B. Das, N. C. Krishnan, B. L. Thomas, and D. J. Cook, "Simple and complex activity recognition through smart phones," in *Proc. Int. Conf. Intell. Environ.*, 2012, pp. 214–221.

[6] M. Ermers, J. Pärkkä, J. Mäntyjärvi, and I. Korhonen, "Detection of daily activities and sports with wearable sensors in controlled and uncontrolled conditions," *IEEE Trans. Inf. Technol. Biomed.*, vol. 12, no. 1, pp. 20–26, Jan. 2008.

[7] J. Farringdon, A. J. Moore, N. Tilbury, J. Church, and P. D. Biemond, "Wearable sensor badge and sensor jacket for context awareness," in *Proc. 3rd Int. Symp. Wearable Comput.*, San Francisco, CA, USA, 1999, pp. 107–113.

[8] D. Graham, G. Simmons, D. T. Nguyen, and G. Zhou, "A software-based sonar ranging sensor for smart phones," *IEEE Internet Things J.*, vol. 2, no. 6, pp. 479–489, Dec. 2015.

[9] A. Graves and J. Schmidhuber, "Offline handwriting recognition with multidimensional recurrent neural networks," in *Proc. Adv. Neural Inf. Process. Syst.*, 2009, pp. 545–552.

[10] A. Graves *et al.*, "A novel connectionist system for unconstrained handwriting recognition," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 31, no. 5, pp. 855–868, May 2009.

[11] J. B. Gomes, S. Krishnaswamy, M. M. Gaber, P. A. C. Sousa, and E. Menasalvas, *Mobile Activity Recognition Using Ubiquitous Data Stream Mining*. Heidelberg, Germany: Springer, 2012, pp. 130–141.

[12] T. Guerreiro, R. Gamboa, and J. Jorge, "Mnemonical body shortcuts: Improving mobile interaction," in *Proc. 15th Eur. Conf. Cognitive Ergonom. Ergonom. Cool Interact.*, 2008, pp. 11–18.

[13] N. Gyorbiro, A. Fabian, and G. Homanyi, "An activity recognition system for mobile phones," *Mobile Netw. Appl.*, vol. 14, no. 1, pp. 82–91, 2009.

[14] B. Hammer, "On the approximation capability of recurrent neural networks," *Neurocomputing*, vol. 31, nos. 1–4, pp. 107–123, 2000.

[15] Z. He, Z. Liu, L. Jin, L.-X. Zhen, and J.-C. Huang, "Weightlessness feature—A novel feature for single tri-axial accelerometer based activity recognition," in *Proc. ICPR*, Tampa, FL, USA, 2008, pp. 1–4.

[16] J. J. C. Ho, "Interruptions: Using activity transitions to trigger proactive messages," Ph.D. dissertation, Dept. Elect. Eng. Comput. Sci., Massachusetts Inst. Technol., Cambridge, MA, USA, 2004.

[17] S. Hochreiter and J. Schmidhuber, "Long short-term memory," *Neural Comput.,* vol. 9, no. 8, pp. 1735–1780, 1997.

[18] R. Jain, "Toward social life networks," *IEEE Comput.*, vol. 47, no. 11, pp. 86–88, Nov. 2014.

[19] R. Jain, L. Jalali, S. Pongpaichet, and A. Gupta, "Building social life networks," *IEEE Data Eng. Bull.*, vol. 36, no. 3, pp. 91–98, Sep. 2013.

[20] N. Kern, B. Schiele, and A. Schmidt, "Recognizing context for annotating a live life recording," *Pers. Ubiquitous Comput.*, vol. 11, no. 4, pp. 251–263, 2007.

[21] J. Lester, T. Choudhury, and G. Borriello, "A practical approach to recognizing physical activities," in *Pervasive Computing*. Heidelberg, Germany: Springer, 2006, pp. 1–16.

[22] D. Liang, Z. Zhang, and M. Peng, "Access point reselection and adaptive cluster splitting-based indoor localization in wireless local area networks," *IEEE Internet Things J.*, vol. 2, no. 6, pp. 573–585, Dec. 2015.

[23] L. Lin, C. Chen, M.-L. Shyu, and S.-C. Chen, "Weighted subspace filtering and ranking algorithms for video concept retrieval," *IEEE Multimedia*, vol. 18, no. 3, pp. 32–43, Mar. 2011.

[24] T. Liu and D. Tao, "Classification with noisy labels by importance reweighting," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 38, no. 3, pp. 447–461, Mar. 2016.

[25] A. Lombardo, C. Panarello, and G. Schembra, "A model-assisted cross-layer design of an energy-efficient mobile video cloud," *IEEE Trans. Multimedia*, vol. 16, no. 8, pp. 2307–2322, Dec. 2014.

[26] A. Mannini, S. S. Intille, M. Rosenberger, A. M. Sabatini, and W. Haskell, "Activity recognition using a single accelerometer placed at the wrist or ankle," *Med. Sci. Sports Exerc.* vol. 45, no. 11, pp. 2193–2203, 2013.

[27] U. Maurer, A. Smailagic, D. P. Siewiorek, and M. Deisher, "Activity recognition and monitoring using multiple sensors on different body positions," in *Proc. Int. Workshop Wearable Implantable Body Sensor Netw.*, Cambridge, MA, USA, 2006, pp. 113–116.

[28] C. McCall, K. Reddy, and M. Shah, "Macro-class selection for hierarchical k-NN classification of inertial sensor data," in *Proc. Int. Conf. Pervasive Embedded Comput. Commun. Syst.*, Rome, Italy, 2012, pp. 106–114.

[29] B. Nham, K. Siangliulue, and S. Yeung, "Predicting mode of transport from iPhone accelerometer data," Mach. Learn. Final Projects, Stanford Univ., Stanford, CA, USA 2008.

[30] S. Pirttikangas, K. Fujinami, and T. Nakajima, "Feature selection and activity recognition from wearable sensors," in *Ubiquitous Computing Systems*. Heidelberg, Germany: Springer, 2006, pp. 516–527.

[31] N. Ravi, N. Dandekar, P. Mysore, and M. L. Littman, "Activity recognition from accelerometer data," in *Proc. Nat. Conf. Artif. Intell.*, 2005, pp. 1541–1546.

[32] S. Samanta and B. Chanda, "Space-time facet model for human activity classification," *IEEE Trans. Multimedia*, vol. 16, no. 6, pp. 1525–1535, Oct. 2014.

[33] H. Si, Y. Kawahara, H. Kurasawa, H. Morikawa, and T. Aoyama, "A context-aware collaborative filtering algorithm for real world oriented content delivery service," in *Proc. ubiPCMM*, 2005.

[34] J. Suutala, S. Pirttikangas, and J. Roning, "Discriminative temporal smoothing for activity recognition from wearable sensors," in *Proc. Ubiquitous Comput. Syst.*, Tokyo, Japan, 2007, pp. 182–195.

[35] D. Tao, L. Jin, Y. Yuan, and Y. Xue, "Ensemble manifold rank preserving for acceleration-based human activity recognition," *IEEE Trans. Neural Netw. Learn. Syst.*, to be published, doi: 10.1109/TNNLS.2014.2357794.

[36] D. Tao, L. Liang, L. Jin, and Y. Gao, "Similar handwritten Chinese character recognition by kernel discriminative locality alignment," *Pattern Recognit. Lett.*, vol. 35, pp. 186–194, Jan. 2014.

[37] M. Thejaswini, P. Rajalakshmi, and U. B. Desai, "Novel sampling algorithm for human mobility-based mobile phone sensing," *IEEE Internet Things J.*, vol. 2, no. 3, pp. 210–220, Jun. 2015.

[38] W. Ugulino *et al.*, "Wearable computing: Accelerometers' data classification of body postures and movements," in *Advances in Artificial Intelligence-SBIA*. Heidelberg, Germany: Springer, 2012, pp. 52–61.

[39] C. Wang, Z. Liu, and S.-C. Chan, "Superpixel-based hand gesture recognition with kinect depth camera," *IEEE Trans. Multimedia*, vol. 17, no. 1, pp. 29–39, Jan. 2015.

[40] M. Weber, G. Bleser, M. Liwicki, and D. Stricker, "Unsupervised motion pattern learning for motion segmentation," in *Proc. ICPR*, Tsukuba, Japan, 2012, pp. 202–205.

[41] J. Xiao, Z. Zeng, and A. Wu, "New criteria for exponential stability of delayed recurrent neural networks," *Neurocomputing*, vol. 134, pp. 182–188, Jun. 2014.

[42] C. Xu, D. Tao, and C. Xu, "Large-margin multi-view information bottleneck," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 36, no. 8, pp. 1559–1572, Aug. 2014.

[43] Y. Yang, H.-Y. Ha, F. C. Fleites, and S.-C. Chen, "A multimedia semantic retrieval mobile system based on HCFGs," *IEEE Multimedia*, vol. 21, no. 1, pp. 36–46, Jan./Mar. 2014.

[44] Z. Yuan, J. Sang, C. Xu, and Y. Liu, "A unified framework of latent feature learning in social media," *IEEE Trans. Multimedia*, vol. 16, no. 6, pp. 1624–1635, Oct. 2014.

[45] O. Yurur, C. H. Liu, and W. Moreno, "Light-weight online unsupervised posture detection by smartphone accelerometer," *IEEE Internet Things J.*, vol. 2, no. 4, pp. 329–339, Aug. 2015.

[46] X. Zhang, Y. Tian, T. Huang, S. Dong, and W. Gao, "Optimizing the hierarchical prediction and coding in HEVC for surveillance and conference videos with background modeling," *IEEE Trans. Image Process.*, vol. 23, no. 10, pp. 4511–4526, Oct. 2014.

[47] X. Zhang, T. Huang, Y. Tian, and W. Gao, "Background-modeling-based adaptive prediction for surveillance video coding," *IEEE Trans. Image Process.*, vol. 23, no. 2, pp. 769–784, Feb. 2014.

[48] L. Zhou and H. Wang, "Toward blind scheduling in mobile media cloud: Fairness, simplicity, and asymptotic optimality," *IEEE Trans. Multimedia*, vol. 15, no. 4, pp. 735–746, Jun. 2013.

[49] L. Zhou, H. Wang, and M. Guizani, "How mobility impacts video streaming over multi-hop wireless networks?" *IEEE Trans. Commun.*, vol. 60, no. 7, pp. 2017–2028, Jul. 2012.

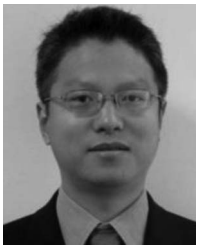
[50] Z. Zeng, T. Huang, and W. X. Zheng, "Multistability of recurrent neural networks with time-varying delays and the piecewise linear activation function," *IEEE Trans. Neural Netw.*, vol. 21, no. 8, pp. 1371–1377, Aug. 2010.

[51] Y. Zheng, W.-K. Wong, X. Guan, and S. Trost, "Physical activity recognition from accelerometer data using a multi-scale ensemble method," in *Proc. 25th Conf. Innov. Appl. Artif. Intell.*, Jul. 2013, pp. 14–18.

Dapeng Tao received the B.E. degree from Northwestern Polytechnical University, Xi'an, China, and the Ph.D. degree from the South China University of Technology, Guangzhou, China.

He is currently a Full Professor with the School of Information Science and Engineering, Yunnan University, Kunming, China. He has authored or coauthored over 30 scientific papers. His current research interests include machine learning, computer vision, and robotics.

Dr. Tao has served over ten international journals, including the IEEE TRANSACTIONS ON NEURAL NETWORKS AND LEARNING SYSTEMS, the IEEE TRANSACTIONS ON MULTIMEDIA, the IEEE TRANSACTIONS ON CIRCUITS AND SYSTEMS FOR VIDEO TECHNOLOGY, the IEEE SIGNAL PROCESSING LETTERS, and *Information Sciences*.



Yonggang Wen (S'99–M'08–SM'14) received the Ph.D. degree in electrical engineering and computer science (with a minor in western literature) from the Massachusetts Institute of Technology, Cambridge, MA, USA, in 2008.

He is currently an Associate Professor with the School of Computer Science and Engineering, Nanyang Technological University, Singapore. He has been with Cisco, San Jose, CA, USA, where he led product development in content delivery network, which had a revenue impact of $3 billion globally. His work in multiscreen cloud social TV has been featured by global media (over 1600 news articles from over 29 countries). He has authored or coauthored over 140 papers in top journals and prestigious conferences. His current research interests include cloud computing, green data center, big data analytics, multimedia networks, and mobile computing.

Dr. Wen serves on Editorial Boards for the IEEE TRANSACTIONS ON CIRCUITS AND SYSTEMS FOR VIDEO TECHNOLOGY, *IEEE Wireless Communication Magazine*, IEEE COMMUNICATIONS SURVEY & TUTORIALS, the IEEE TRANSACTIONS ON MULTIMEDIA, the IEEE TRANSACTIONS ON SIGNAL AND INFORMATION PROCESSING OVER NETWORKS, IEEE ACCESS, and *Elsevier Ad hoc Networks*. He was elected as the Chair for the IEEE ComSoc Multimedia Communication Technical Committee from 2014 to 2016. He was a recipient of the ASEAN ICT Award 2013 (Gold Medal) and the Data Centre Dynamics Awards 2015—APAC, for his work on cloud 3-D view, as the only academia entry and a corecipient of the 2015 IEEE Multimedia Best Paper Award, and the Best Paper Awards at EAI/ICST Chinacom 2015, IEEE WCSP 2014, IEEE Globecom 2013, and IEEE EUC 2012.

Richang Hong received the Ph.D. degree from the University of Science and Technology of China, Hefei, China, in 2008.

He is currently a Professor with the Hefei University of Technology, Hefei, China. From 2008 to 2010, he was a Research Fellow with the School of Computing, National University of Singapore. He has coauthored over 70 publications in the areas of his research interests, which include multimedia content analysis and social media.

Dr. Hong served as the Associate Editor of *Information Sciences and Signal Processing* (Elsevier) and the Technical Program Chair of MMM 2016. He is a Member of ACM and an Executive Committee Member of the ACM SIGMM China Chapter. He was a recipient of the Best Paper Award of ACM Multimedia 2010, the Best Paper Award of ACM ICMR 2015, and the Honorable Mention of the IEEE TRANSACTIONS ON MULTIMEDIA Best Paper Award.