Human Daily Activity Recognition With Sparse Representation Using Wearable Sensors

Mi Zhang, *Student Member, IEEE*, and Alexander A. Sawchuk, *Life Fellow, IEEE*

***Abstract*—Human daily activity recognition using mobile personal sensing technology plays a central role in the field of pervasive healthcare. One major challenge lies in the inherent complexity of human body movements and the variety of styles when people perform a certain activity. To tackle this problem, in this paper, we present a novel human activity recognition framework based on recently developed compressed sensing and sparse representation theory using wearable inertial sensors. Our approach represents human activity signals as a sparse linear combination of activity signals from all activity classes in the training set. The class membership of the activity signal is determined by solving a ℓ^1 minimization problem. We experimentally validate the effectiveness of our sparse representation-based approach by recognizing nine most common human daily activities performed by 14 subjects. Our approach achieves a maximum recognition rate of 96.1%, which beats conventional methods based on nearest neighbor, naive Bayes, and support vector machine by as much as 6.7%. Furthermore, we demonstrate that by using random projection, the task of looking for "optimal features" to achieve the best activity recognition performance is less important within our framework.**

***Index Terms*—Compressed sensing, human activity recognition, pervasive healthcare, sparse representation, wearable computing.**

I. Introduction

PERVASIVE healthcare, the use of pervasive computing technologies to integrate healthcare monitoring and intervention seamlessly into people's everyday lives, is a rapidly expanding area of research that attracts more and more researchers in recent years. As one of the key problems within its domain, automatically recognizing and logging human daily activities has significant potential to increase the efficiency of healthcare providers [1]. Among all the human activity sensing technologies, wearable sensing system has the advantage of being with people throughout the day, enabling continuously collecting people's activity information. This unique characteristic makes it a perfect platform to deliver long-term personalized healthcare anywhere and any time, enabling a wide range of healthcare applications including personal fitness monitoring, preventive and chronic healthcare, and elderly support [2].

As a pattern recognition problem, recognizing human daily activities automatically is challenging due to the complex nature of human body movements. Over the years, many research studies have been done to analyze both simple and complex human activities based on wearable sensors. A large number of them focus on studying what are the most informative features that can be extracted from the activity data and what is the most effective learning machine for classifying these activities [3]–[5]. Another line of research tries to answer the question of what is the most appropriate computational model to represent human activity data [6]–[8]. However, despite the significant research efforts, the scalability of handling large intraclass variations and the robustness to model parameters of many existing human activity recognition techniques are still quite limited.

During the past few years, research on high-dimensional sparse signals has experienced great breakthroughs. A sparse signal is "*a signal that can be represented as a linear combination of relatively few base elements in a basis or an overcomplete dictionary*" [9]. In fact, many real-world signals are high dimensional and sparse in nature. For instance, smooth images exhibit sparse structures if represented using a Fourier basis. Similarly, piecewise smooth images can be regarded as sparse signals under a wavelet basis [10]. Although the original goal of exploring a signal's sparsity is for signal compression and reconstruction, its discriminative nature has been exploited and successfully applied to many machine learning/pattern recognition tasks. Examples include human face recognition [11], iris identification [12], facial action unit recognition [13], human speech recognition [14], and object recognition [15]. Inspired by their success based on their highly scalable ways of data modeling, in this paper, we explore the sparsity nature of human daily activity signals sampled from wearable sensors and propose a sparse representation-based framework for human daily activity modeling and recognition. An important step of our approach is the selection of a basis or the design of the overcomplete dictionary for sparse representation. One option is to use the standard sparsity-inducing bases such as Fourier basis, wavelets, and curvelets used in many image processing techniques. In this study, we follow the method proposed in [11] to use the training samples directly as the basis to construct the overcomplete dictionary. The rationale behind this strategy is that if sufficient training samples are available, it is assumed that a test sample can be well represented by a linear combination of training examples from the same activity class, and therefore, the representation of the test sample in terms of the training samples is naturally sparse. We have adopted this framework in our previous work to recognize the location of the wearable sensors on human bodies [16]. The main goal of this paper is to study the key characteristics of the sparse representation-based framework for human activity modeling and recognition.

Manuscript received May 25, 2012; revised October 14, 2012 and January 14, 2013; accepted March 11, 2013. Date of publication March 20, 2013; date of current version May 1, 2013.

The authors are with the Ming Hsieh Department of Electrical Engineering, University of Southern California, Los Angeles, CA 90089 USA (e-mail: mizhang@usc.edu; sawchuk@sipi.usc.edu).

Color versions of one or more of the figures in this paper are available online at http://ieeexplore.ieee.org.

Digital Object Identifier 10.1109/JBHI.2013.2253613

Fig. 1. MotionNode sensor and its placement. (a) MotionNode platform. (b) During data collection, a single MotionNode is packed firmly into a mobile phone pouch and attached to the participant's front right hip.

To achieve this goal, we study the robustness of the framework related to different feature dimensions and different selections of features. We also compare the recognition performance obtained with sparse representation to some conventional activity recognition approaches such that the advantages of the proposed sparse representation-based approach can be clearly and better illustrated.

The organization of the paper is as follows. We first give a brief review on some of the existing approaches for human activity recognition using wearable sensors. Then, we introduce the sensing platform and dataset used for this study. Next, we describe the details of our sparse representation-based framework and present the experimental results. Finally, we conclude this paper and outline our future work.

II. Related Work

A large number of existing human activity recognition techniques concentrate on feature studies. For example, in [17], the authors extracted features based on linear-discriminant analysis (LDA) and applied artificial neural nets for classification. Their procedure achieved 97.9% average recognition accuracy over 15 different human activities and three static states. In [18], the authors claimed that the classification performance could be improved by selecting features for each activity separately. Similar results were also found in [19] where the authors used feature selection techniques to demonstrate that a feature's usefulness was dependent on the specific activity to be inferred.

Another line of research focuses on designing computational models to represent human activity signals. In [7], the authors used motion trajectories in 3-D space to characterize each activity. In [20], the authors proposed to use motion primitives to build activity models where motion primitives are defined as the most basic units shared by all the activity classes. Recently, researchers have attempted to model activities using manifold learning, in which the goal is to capture the intrinsic low-dimensional manifold structure embedded in the activity signals. For example, the authors in [8] used local linear embedding (LLE) to capture the geometric structure of the activity signals. As another example, the authors in [21] applied isomap to construct the low-dimensional activity manifolds and used a dynamic time warping technique for manifold recognition.

Unlike the existing approaches just mentioned, the framework we propose in this study leverages the sparsity property of human activity signals and builds the human activity models from a totally different perspective. As we demonstrate in the following sections, the sparse representation-based approach has great potential in handling the feature sensitivity issue reported in both [18] and [19].

TABLE I
Statistics of the Participants for Data Collection

	Age	Height (cm)	Weight (kg)
range	21 - 49	160 - 185	43 - 80
mean	30.1	170	64.6
std	7.2	6.8	12.1

III. Sensing Platform and Dataset

The wearable sensing device we use for this study is called MotionNode[1] [see Fig. 1(a)]. MotionNode is an accurate high-performance inertial measurement unit that integrates a three-axis accelerometer, a three-axis gyroscope, and a three-axis magnetometer. It has been widely used in real-time 3-D motion tracking applications. In this study, we use the three-axis accelerometer and the three-axis gyroscope to sense the human activity signals. Each axis of the accelerometer and the gyroscope can sense ± 6 g acceleration and ± 500 dps rotation rate, respectively. The sampling rates for both accelerometer and gyroscope are set to 100 Hz. This is high enough to capture all the details of the person's movements. Another important reason we use MotionNode is that its sensor size is extremely small such that it can be attached to a person's body in a nonintrusive manner.

Fourteen participants with diversity in gender (7 male, 7 female), age, height, and weight were recruited for this study. The statistics of age, height, and weight are listed in Table I. The data collection was carried out in an open environment. Each participant was asked to perform nine types of activities: *walk forward*, *walk left*, *walk right*, *go upstairs*, *go downstairs*, *jump up*, *run*, *stand*, and *sit*. We chose these activities because they correspond to the most basic and common activities in people's daily lives and are useful for applications such as physical fitness monitoring and assisted living for elderly people. During data collection, a MotionNode was packed inside a mobile phone pouch that was attached firmly onto the participant's right front hip. A miniature laptop was used to record sampled data from MotionNode [see Fig. 1(b)]. To capture day-to-day variations and reduce the interindividual correlation to the minimum, each participant was asked to perform five trials for each activity on different days at various indoor and outdoor locations with no instructions on how the activities should be performed. By doing this, participants could perform these activities based on their own styles. We expect that the diversity in performance style could cover a wide range of population.

[1] http://www.motionnode.com/

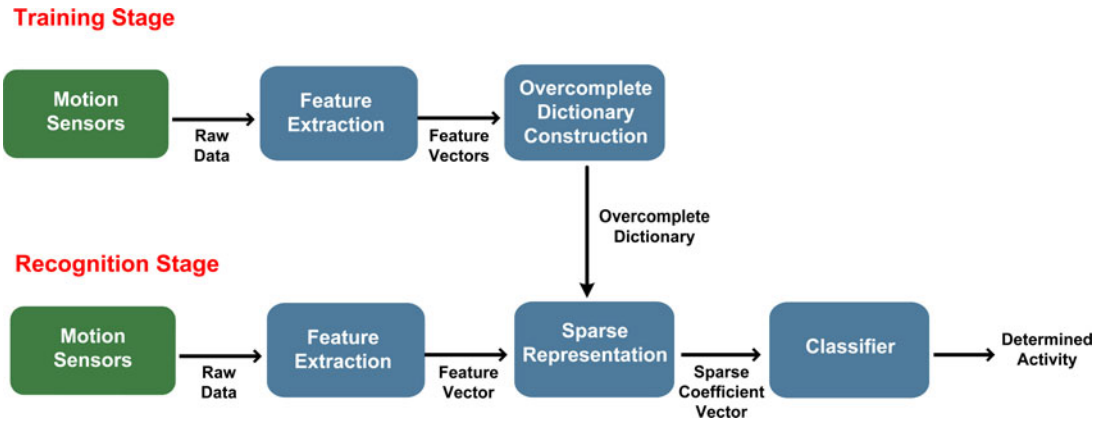

Fig. 2. Block diagram of the sparse representation-based human activity recognition framework.

TABLE II
FEATURES CONSIDERED IN THIS STUDY

Mean	Median	Standard Deviation
Variance	Root Mean Square	Interquartile Range
First Order Derivative	Second Order Derivative	Skewness
Kurtosis	Zero Crossing Rate	Mean Crossing Rate
Pairwise Correlation	Energy	Dominant frequency
Spectral Entropy	Movement Intensity	Signal Magnitude Area
Acceleration Correlation	Averaged Acceleration Energy	Averaged Velocity
Eigenvalues of Dominant Directions	Averaged Rotation Angles	Averaged Rotation Energy

IV. SPARSE REPRESENTATION-BASED FRAMEWORK

The block diagram of our sparse representation-based framework for human activity modeling and recognition is illustrated in Fig. 2. The proposed framework consists of two stages: training stage and recognition stage. In the training stage, a sliding window is used to segment the streaming activity signals sampled from the wearable sensors into a sequence of fixed-length windows. In this study, the window is set to be 4 s long with 50% overlap. Here, we assume that all the important information of each activity is contained inside each window. This information is extracted by computing various features over the sampled sensor data within each window to form a feature vector. The feature vectors from training samples of all activity classes are then concatenated together to construct the overcomplete dictionary. In the recognition stage, the unknown stream of activity signal is first segmented and transformed into a feature vector in the same manner as in the training stage. Its sparse representation based on the overcomplete dictionary constructed in the training stage is then extracted and imported into the classifier for classification. In the remainder of this section, we explain every component of this framework in detail.

A. Feature Extraction

For wearable sensor-based human activity recognition, a variety of features have been intensively investigated in previous studies. Table II shows the list of features we consider in this study. Some of them, such as mean, variance, correlation, and entropy, have proved to be useful for wearable sensor-based human activity recognition. Features like zero crossing rate, mean crossing rate, and first-order derivative are also considered because these features have been successfully applied in similar recognition problems such as human speech recognition and handwriting recognition problems. We also include physical features proposed in [5] such as movement intensity, eigenvalues of dominant directions, and averaged velocity along heading direction since they have been reported to be able to boost the classification performance. All these features are extracted from both accelerometer and gyroscope. In total, the dimensionality of the input feature space is 110.

B. Feature Selection Versus Random Projection

As stated in Section I, one goal of this work is to study how the choices of features and the dimensions of feature space affect the performance of our sparse representation-based framework. We want to understand not only the effect of using the features in Table II themselves but also the effect of using the linear combinations of these features based on some linear transformations. There are many types of linear transformations that can be used. Popular examples include principal component analysis (PCA) [22] and LDA [23]. In this study, we are particularly interested in using random projection as the linear transformation since it has been proved very powerful to encode information effectively in many applications and can be implemented much more efficiently than PCA and LDA [24]. In random projection, a linear transformation is represented as a random matrix R whose entries are independent and identically distributed (i.i.d.) random variables from a zero-mean Gaussian distribution in which each row is normalized to have unit length [25]. As a result, the newly generated features are linear combinations of all the original features in Table II with randomly generated coefficients. Compared to the features generated by PCA and LDA, these randomly projected features are less structured but encode the information from all the original features. We will compare the effects of using the original features selected from the feature set listed in Table II and the transformed features generated by

random projection on the recognition performance in the next section.

C. Overcomplete Dictionary Construction and Sparse Representation

Assume that there are k distinct activity classes to recognize and n_i training samples from class i, $i \in [1, 2, \ldots, k]$. Recalling that each training sample is represented as an m-dimensional feature vector, we arrange the given n_i training samples from class i as columns of a data matrix $\boldsymbol{D}_i = [\boldsymbol{x}_{i,1}, \boldsymbol{x}_{i,2}, \ldots, \boldsymbol{x}_{i,n_i}] \in \mathbb{R}^{m \times n_i}$. Here, we make a key assumption that given sufficient training samples from class i (i.e., the number of columns of the data matrix $\boldsymbol{D}_i$ is large enough), any new test sample $\boldsymbol{y} \in \mathbb{R}^m$ that belongs to the same activity class can be approximately represented as a linear combination of the training samples in $\boldsymbol{D}_i$

$$\boldsymbol{y} = \alpha_{i,1}\boldsymbol{x}_{i,1} + \alpha_{i,2}\boldsymbol{x}_{i,2} + \cdots + \alpha_{i,n_i}\boldsymbol{x}_{i,n_i} \tag{1}$$

with coefficients $\alpha_{i,j} \in \mathbb{R}, j = 1, 2, \ldots, n_i$.

Next, we define a new matrix $\boldsymbol{A}$ that concatenates the training samples from all the activity classes as

$$\begin{aligned}\boldsymbol{A} &= [\boldsymbol{D}_1, \boldsymbol{D}_2, \ldots, \boldsymbol{D}_k] \in \mathbb{R}^{m \times n} \\ &= [\boldsymbol{x}_{1,1}, \ldots, \boldsymbol{x}_{1,n_1}, \boldsymbol{x}_{2,1}, \ldots, \boldsymbol{x}_{2,n_2}, \ldots, \boldsymbol{x}_{k,1}, \ldots, \boldsymbol{x}_{k,n_k}]\end{aligned} \tag{2}$$

where $n = n_1 + n_2 + \cdots + n_k$. In such case, the matrix $\boldsymbol{A}$ can be seen as an overcomplete dictionary of n prototype elements. Based on the assumption just mentioned, we can express the test sample $\boldsymbol{y}$ from class i in terms of the overcomplete dictionary $\boldsymbol{A}$ as

$$\boldsymbol{y} = \boldsymbol{A}\boldsymbol{\alpha} \tag{3}$$

where

$$\boldsymbol{\alpha} = [0, \ldots, 0, \alpha_{i,1}, \alpha_{i,2}, \ldots, \alpha_{i,n_i}, 0, \ldots, 0]^T \tag{4}$$

is a sparse coefficient vector whose entries are zero except those associated with class i. Therefore, $\boldsymbol{\alpha}$ can be regarded as a sparse representation of $\boldsymbol{y}$ based on the overcomplete dictionary $\boldsymbol{A}$. More importantly, the entries of $\boldsymbol{\alpha}$ encode the identity of $\boldsymbol{y}$. In other words, we can infer the class membership of the test sample $\boldsymbol{y}$ by finding the solution of the linear system of equations of (3) that is expected to be sparse.

D. Sparse Recovery via ℓ^1 Minimization

Based on the theory of linear algebra, the solution of the linear system of equations of (3) depends on the characteristic of the matrix $\boldsymbol{A}$. If $m > n$, the system $\boldsymbol{y} = \boldsymbol{A}\boldsymbol{\alpha}$ is overdetermined and the solution can be found uniquely. However, in most real-world applications, the number of prototypes in the overcomplete dictionary is typically much larger than the dimensionality of the feature representation (i.e., $m \ll n$). Therefore, the linear system of equations of (3) is underdetermined and has no unique solution.

Traditionally, this difficulty is solved by choosing the minimum ℓ^2 solution. That is, the desired coefficients have minimum ℓ^2 norm

$$\hat{\boldsymbol{\alpha}} = \arg\min_{\boldsymbol{\alpha}} \|\boldsymbol{\alpha}\|_2 \text{ subject to } \boldsymbol{y} = \boldsymbol{A}\boldsymbol{\alpha} \tag{5}$$

where $\|.\|_2$ denotes the ℓ^2 norm. The solution of the aforementioned problem is given by

$$\hat{\boldsymbol{\alpha}} = \boldsymbol{A}^{\dagger}\boldsymbol{y} \tag{6}$$

where $\boldsymbol{A}^{\dagger}$ is the pseudoinverse of $\boldsymbol{A}$. However, this solution yields a nonsparse coefficient vector that is not informative for our activity recognition task.

Recent research in the field of compressed sensing [26], [27] has shown that if $\boldsymbol{\alpha}$ is sufficiently sparse, it can be recovered by solving the ℓ^1 minimization problem instead

$$\hat{\boldsymbol{\alpha}} = \arg\min_{\boldsymbol{\alpha}} \|\boldsymbol{\alpha}\|_1 \text{ subject to } \boldsymbol{y} = \boldsymbol{A}\boldsymbol{\alpha} \tag{7}$$

where $\|.\|_1$ denotes the ℓ^1 norm. This optimization problem, also known as Basis Pursuit (BP), is built on a solid theoretical basis and can be solved very efficiently with traditional linear programming techniques whose computational complexities are polynomial in n [28].

In practical real-world applications, signals are always noisy. As a result, it may not be possible or reasonable to model the test sample exactly as a sparse linear combination of the training samples as in (3). In such cases, (3) can be modified to explicitly account for limited noise with

$$\boldsymbol{y} = \boldsymbol{A}\boldsymbol{\alpha} + \boldsymbol{e} \tag{8}$$

where $\boldsymbol{e}$ is the noise term with bounded energy $\|\boldsymbol{e}\|_2 < \epsilon$. With such modification, the sparse solution of (8) can still be efficiently computed by solving the following ℓ^1 minimization problem via second-order cone programming [28]:

$$\hat{\boldsymbol{\alpha}} = \arg\min_{\boldsymbol{\alpha}} \|\boldsymbol{\alpha}\|_1 \text{ subject to } \|\boldsymbol{A}\boldsymbol{\alpha} - \boldsymbol{y}\|_2 \leq \epsilon. \tag{9}$$

E. Classification via Sparse Representation

Given a new test sample $\boldsymbol{y}$ in the form of an m-dimensional feature vector from one of the k activity classes, we first compute its sparse coefficient vector $\hat{\boldsymbol{\alpha}}$ by solving (9). To identify the class membership of $\boldsymbol{y}$, we adopt the classification strategy by comparing how well the various parts of the coefficient vector $\hat{\boldsymbol{\alpha}}$ associated with different activity classes can reproduce $\boldsymbol{y}$, where the reproduction error is measured by the residual value. Specifically, the residual of class i is defined as

$$r_i(\boldsymbol{y}) = \|\boldsymbol{y} - \boldsymbol{A}\delta_i(\hat{\boldsymbol{\alpha}})\|_2 \tag{10}$$

where $\delta_i(\hat{\boldsymbol{\alpha}})$ is a characteristic function that selects only the coefficients in $\hat{\boldsymbol{\alpha}}$ associated with class i. Therefore, $r_i(\boldsymbol{y})$ measures the difference between the true solution $\boldsymbol{y}$ and the approximation using only the components from class i. Finally, $\boldsymbol{y}$ is classified as the activity class c that gives rise to the smallest residual value

$$c = \arg\min_i r_i(\boldsymbol{y}). \tag{11}$$

As an example, Fig. 3(a) and (c) illustrates the two coefficient vectors recovered by solving (9) with the noise tolerance $\epsilon = 0.03$ for two test samples from two activities: *walk forward* and *running*, respectively. As illustrated, both of the recovered

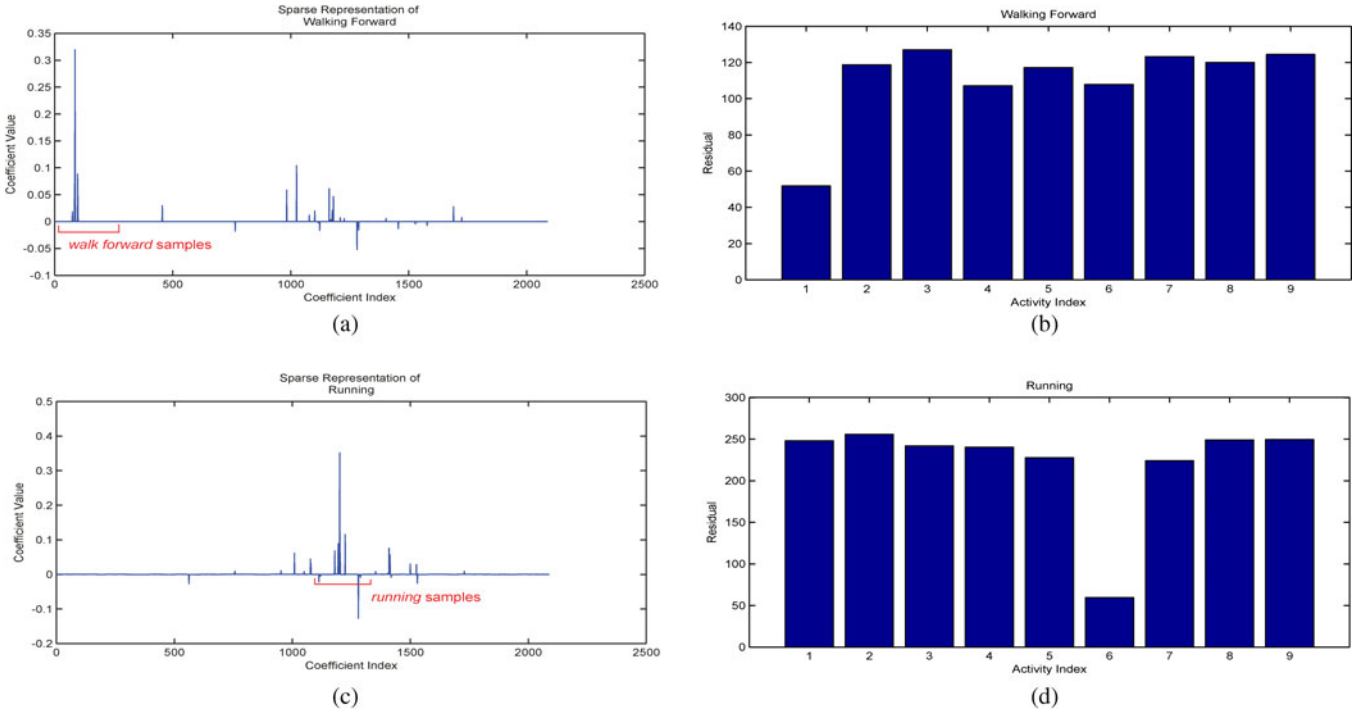

Fig. 3. Sparse representation solutions via ℓ^1 minimization and the corresponding residuals for two test samples from *walk forward* and *running*, respectively. (a) Sparse coefficient solution recovered via ℓ^1 minimization for one test sample from activity *walk forward*. (b) Residual values with respect to the nine activity classes. The test sample is correctly classified as *walk forward* (index number 1). The ratio between the two smallest residuals is 1:2.2. (c) Sparse coefficient solution recovered via ℓ^1 minimization for one test sample from activity *running*. (d) Residual values with respect to the nine activity classes. The test sample is correctly classified as *running* (index number 6). The ratio between the two smallest residuals is 1:3.8.

coefficient vectors are sparse. Moreover, the majority of the large coefficients are associated with the training samples belonging to the same activity class. Fig. 3(b) and (d) shows the corresponding residual values with respect to the nine activity classes. As illustrated, both test samples are correctly classified since the smallest residual value is associated with the true activity class. To show the robustness of our residual-based classification strategy, we calculate the ratios between the two smallest residuals for each test sample. The larger the ratio value, the more robust the classification result. In the examples of Fig. 3(b) and (d), the ratios between the two smallest residuals are 1:2.2 for *walk forward* and 1:3.8 for *running*. We obtain similar results for the other seven activities.

F. Classification Confidence Measure

The residual-based classification strategy described in the previous section only provides a classification result whose confidence is unknown. To quantify the classification confidence, we use the metric sparsity concentration index (SCI) proposed in [11]. The rationale behind the design of SCI is based on the sparse recovery results illustrated in Fig. 3. Specifically, a test sample classified with high confidence should have a sparse representation whose nonzero entries concentrate mostly on one activity class, whereas a test sample classified with low confidence should have sparse coefficients spread widely among multiple activity classes.

Based on this observation, the SCI of a coefficient vector $\boldsymbol{\alpha}$ is defined as

$$\mathrm{SCI}(\boldsymbol{\alpha}) = \frac{k \cdot \max_i \|\delta_i(\boldsymbol{\alpha})\|_1 / \|\boldsymbol{\alpha}\|_1 - 1}{k-1} \tag{12}$$

With this definition, SCI takes values between 0 and 1. For a coefficient vector of a test sample recovered via ℓ^1 minimization, if the SCI value is close to 1, that is, the value of $\max_i \|\delta_i(\boldsymbol{\alpha})\|_1 / \|\boldsymbol{\alpha}\|_1 \approx 1$, it indicates that the test sample can be approximately represented using training samples only from a single activity class. On the other hand, if the SCI value is close to 0, that is, $\max_i \|\delta_i(\boldsymbol{\alpha})\|_1 / \|\boldsymbol{\alpha}\|_1 \approx 1/k$, it corresponds to the situation where no single activity is dominant such that the nonzero coefficients are distributed over all activity classes. For example, the SCI values of the two test samples from *walk forward* and *running* in Fig. 3 are 0.44 and 0.57, respectively. Furthermore, we can manually set up a threshold $\tau \in [0, 1]$ such that only test samples with SCI values equal or larger than τ are considered. As will be shown in the next section, τ can be used as an input parameter of the overall human activity recognition system such that it can be tuned by users to achieve the desired performance.

V. Experiments and Results

In this section, we evaluate the performance of our sparse representation-based framework. For the evaluation procedure, we use the leave-one-subject-out cross-validation strategy.

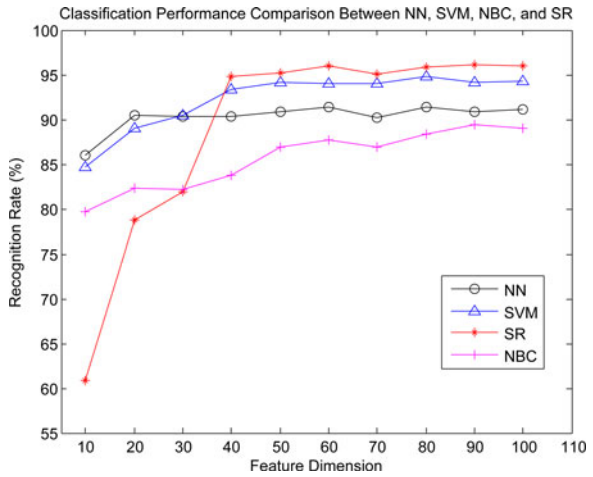

Fig. 4. Impact of feature dimension.

Specifically, we use the data from 13 subjects as training examples. Data from the left-out subject are used for testing. This process iterates for every subject. The final result is the averaged value over all the subjects.

The framework was implemented in MATLAB programming environment. The ℓ^1 minimization was performed using the ℓ^1-magic package.[2] The noise tolerance ϵ was set to 0.03. We use the classification accuracy as the single quality metric for all our experiments. The classification accuracy *ACC* is defined as

$$ACC = \frac{TP + TN}{TP + TN + FP + FN} \tag{13}$$

where the variables *TP*, *TN*, *FP*, and *FN*, respectively, represent the number of *True Positive*, *True Negative*, *False Positive*, and *False Negative* outcomes in a given experiment.

A. Effect of the Feature Dimension and Comparison to Baseline Algorithms

As our first experiment, we examine the framework's classification performance with respect to different feature dimensions, and compare it to three classical classification methods: nearest neighbor (NN), naive Bayesian classifier (NBC), and support vector machine (SVM). We choose these three classification methods as the baseline algorithms because the advantages of our proposed sparse representation-based approach can be clearly and better illustrated. In particular, Fig. 4 illustrates the average classification accuracy rates as a function of feature dimension, ranging from 10 to 100, with interval equal to 10. Each curve represents one classification method, respectively. At each dimension, features are selected based on the sequential forward selection (SFS) since it is reported in [5] as a very effective feature selection method. In such case, we compare the performance of the four classification methods using the same feature set.

As shown in the figure, NN and NBC have relatively worse performance. The maximum recognition rates for NN and NBC are 91.3% and 89.4%, respectively. In comparison, SVM achieves a better performance than both NN and NBC at each individual feature dimension when feature dimension is equal or larger than 30, with the best rate of 94.8%. For our sparse representation-based classification method (SR), the performance is the worst when feature dimension is less than or equal to 30. This observation indicates that using fewer than 30 features is not sufficient to recover the sparse signals via ℓ^1 minimization with no information loss. However, when the feature dimension is equal to or larger than 40, SR achieves consistent performance and beats all the other three classification methods, achieving a maximum recognition rate of 96.1% when feature dimension is equal to 60.

To take a closer look at the classification result, Table III shows the confusion table for feature dimension equal to 50. The overall averaged recognition accuracy across all activities is 95.2%, with eight out of nine activities having precision and recall values higher than 90%. If we examine the recognition performance for each activity individually, both *walk left* and *walk right* achieve very high precision and recall values. Furthermore, they never get confused with each other. For *jump up*, although it has a near 100% precision value, it only achieves a recall value of 92.9%. This is because some of the samples of *jump up* are misclassified as *run forward* but not vice versa. For *walk forward*, it is interesting to notice that it can be misclassified as any other walk-related activities. This is exactly the same as *go upstairs* except that *go upstairs* never gets misclassified to *go downstairs*. Finally, *sit on a chair* has a relatively low recall value because it is mostly confused with *stand*. This result makes sense since both *sit on a chair* and *stand* are static activities, and we expect difficulty in differentiating different static activities especially when the sensing device is attached to the hip of the participants.

B. Effect of the Choice of Features and Random Projection

In this section, we study the effect of different choices of features and random projection on the classification performance of our framework. Similar to the previous section, Fig. 5 shows the average classification accuracy rates versus feature dimension from 10 to 100. The red curve with asterisks represents the standard SR method using features selected based on the SFS feature selection method mentioned previously. The black curve with circles represents the standard SR method using features randomly selected without the help from any feature selection algorithm. The blue curve with triangles represents the SR method with random projection. In this study, the entries of our random projection matrix are independently sampled from a Gaussian distribution with mean zero and variance $1/n$ (recall n is the total number of training samples included in the overcomplete dictionary). As shown, it is interesting to see that the SR method with random projection achieves very similar performance to the standard SR method using features selected from SFS. In comparison, the SR method with randomly selected features performs much worse than the other two methods when using 80 or fewer features. This observation indicates that the choice of features

[2]http://www-stat.stanford.edu/∼candes/l1magic/

TABLE III
CONFUSION TABLE FOR OUR SPARSE REPRESENTATION-BASED HUMAN ACTIVITY RECOGNITION FRAMEWORK WHEN FEATURE DIMENSION IS 50

Ground Truth \ Classified Activity	Walk forward	Walk left	Walk right	Go up stairs	Go down stairs	Run forward	Jump up	Sit on a chair	Stand	Total	Recall
1 Walk forward	1710	15	14	20	6	2	0	1	2	1770	96.6%
2 Walk left	10	1528	0	2	14	0	0	0	0	1554	98.3%
3 Walk right	29	0	1673	22	1	3	0	6	0	1734	96.5%
4 Go up stairs	31	8	19	1372	0	5	0	0	7	1442	95.1%
5 Go down stairs	8	16	0	5	1352	5	0	0	0	1386	97.5%
6 Run forward	0	0	0	0	0	925	1	0	0	926	99.9%
7 Jump up	0	0	0	0	10	43	692	0	0	745	92.9%
8 Sit on a chair	5	0	2	4	1	0	4	1387	91	1494	92.8%
9 Stand	11	7	0	4	4	0	0	164	1188	1378	86.2%
Total	1804	1574	1708	1429	1388	983	697	1558	1288		
Precision	94.8%	97.1%	98.0%	96.0%	97.4%	94.1%	99.3%	89.0%	92.2%		

The entry in the i^{th} row and j^{th} column is the count of activity instances from class i but classified as class j. Overall classification accuracy is 95.2%.

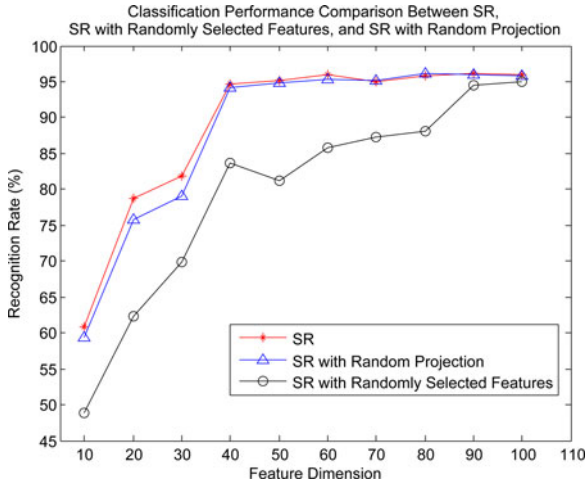

Fig. 5. Impact of feature choices.

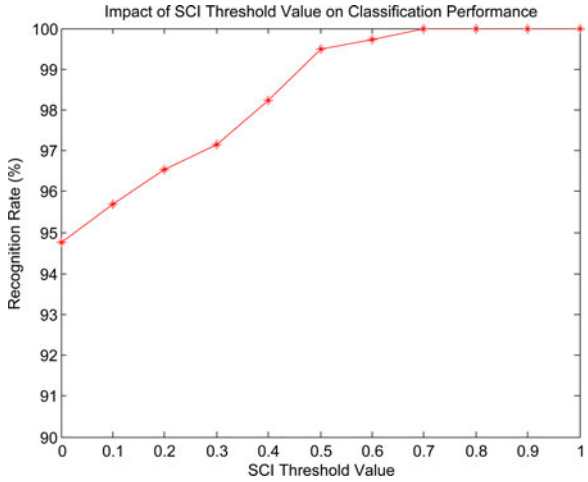

Fig. 6. Impact of SCI threshold value on classification performance.

plays a significant role in our sparsity-based framework if random projection is not used. On the other hand, by using random projection, the task of looking for "optimal features" to achieve the best performance is less important. In other words, the random projected features should perform as well as features selected by many effective feature selection algorithms.

C. SCI as a Measure of Confidence

Finally, as our last experiment, we examine the role of the classification confidence measure SCI as a tunable input parameter of the recognition system.

As explained in the previous section, we set up a threshold $\tau \in [0, 1]$ such that only test samples with SCI values equal or larger than τ are considered. Fig. 6 shows the average classification accuracy rates by sweeping the threshold τ through a range of values starting from 0 to 1, with interval equal to 0.1. As expected, the accuracy rate increases monotonically as the threshold τ increases from 0 to 1. The classification accuracy reaches 100% when the threshold τ is at 0.7. Based on this generated curve, the user can set the threshold τ to a specific value to make the recognition system achieve the desired performance.

VI. CONCLUSION AND FUTURE WORK

In this paper, we presented a sparse representation-based framework for human daily activity modeling and recognition using wearable sensors. In conclusion, our framework achieves a better performance compared to the methods using NN, NBC, and SVM when the feature dimensions are equal to or larger than 40. Furthermore, we demonstrate that by using random projection, our framework can achieve competitive performance compared to the one with finely selected features. Based on the promising results exhibited from our approach, we plan to test it on larger datasets containing more types of human activities. We will also consider implementing this approach on mobile phones and build activity recognition-based healthcare applications for field studies in the near future.

REFERENCES

[1] C. Orwat, A. Graefe, and T. Faulwasser, "Towards pervasive computing in health care—a literature review," *BMC Med. Inf. Decis. Making*, vol. 8, no. 26, pp. 1–19, 2008.

[2] T. Choudhury, S. Consolvo, B. Harrison, J. Hightower, A. LaMarca, L. LeGrand, A. Rahimi, A. Rea, G. Bordello, B. Hemingway, P. Klasnja, K. Koscher, J. Landay, J. Lester, D. Wyatt, and D. Haehnel, "The mobile sensing platform: An embedded activity recognition system," *IEEE Pervasive Comput.*, vol. 7, no. 2, pp. 32–41, Apr. 2008.

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

[4] N. Ravi, N. Dandekar, P. Mysore, and M. L. Littman, "Activity recognition from accelerometer data," in *Proc. Conf. Innov. Appl. Artif. Intell.*, Pittsburgh, PA, USA, 2005, pp. 1541–1546.

[5] M. Zhang and A. A. Sawchuk, "A feature selection-based framework for human activity recognition using wearable multimodal sensors," in *Proc. Int. Conf. Body Area Networks*, Beijing, China, Nov. 2011, pp. 92–98.

[6] P. Lukowicz, J. A. Ward, H. Junker, M. Stager, G. Troster, A. Atrash, and T. Starner, "Recognizing workshop activity using body worn microphones and accelerometers," in *Proc. Int. Conf. Pervasive Comput.*, Vienna, Austria, Apr. 2004, pp. 18–32.

[7] T. Stiefmeier, D. Roggen, and G. Tröster, "Gestures are strings: Efficient online gesture spotting and classification using string matching," in *Proc. Int. Conf. Body Area Networks*, Florence, Italy, Jun. 2007, pp. 16:1–16:8.

[8] M. Zhang and A. A. Sawchuk, "Manifold learning and recognition of human activity using body-area sensors," in *Proc. IEEE Int. Conf. Mach. Learn. Appl.*, Honolulu, HI, USA, Dec. 2011, pp. 7–13.

[9] R. Baraniuk, E. Candes, M. Elad, and Y. Ma, "Applications of sparse representation and compressive sensing [scanning the issue]," *Proc. IEEE*, vol. 98, no. 6, pp. 906–909, Jun. 2010.

[10] M. B. Wakin, J. N. Laska, M. F. Duarte, D. Baron, S. Sarvotham, D. Takhar, K. F. Kelly, and R. G. Baranuik, "Compressive imaging for video representation and coding," presented at the Picture Coding Symp., Beijing, China, Apr. 2006.

[11] J. Wright, A. Y. Yang, A. Ganesh, S. S. Sastry, and Y. Ma, "Robust face recognition via sparse representation," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 31, no. 2, pp. 210–227, Feb. 2009.

[12] J. Pillai, V. Patel, R. Chellappa, and N. Ratha, "Secure and robust iris recognition using random projections and sparse representations," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 33, no. 9, pp. 1877–1893, Sep. 2011.

[13] M. H. Mahoor, M. Zhou, K. L. Veon, S. M. Mavadati, and J. F. Cohn, "Facial action unit recognition with sparse representation," in *Proc. IEEE Int. Conf. Autom. Face Gesture Recognit.*, Santa Barbara, CA, USA, Mar. 2011, pp. 336–342.

[14] J. F. Gemmeke and B. Cranen, "Using sparse representations for missing data imputation in noise robust speech recognition," presented at the Eur. Signal Process. Conf., Lausanne, Switzerland, Aug. 2008.

[15] J. Mairal, F. Bach, J. Ponce, G. Sapiro, and A. Zisserman, "Discriminative learned dictionaries for local image analysis," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Anchorage, AK, USA, Jun. 2008, pp. 1–8.

[16] W. Xu, M. Zhang, A. A. Sawchuk, and M. Sarrafzadeh, "Co-recognition of human activity and sensor location via compressed sensing in wearable body sensor networks," in *Proc. IEEE Int. Conf. Wearable Implantable Body Sensor Netw.*, London, U.K., May 2012, pp. 124–129.

[17] A. Khan, Y.-K. Lee, S. Lee, and T.-S. Kim, "A triaxial accelerometer-based physical-activity recognition via augmented-signal features and a hierarchical recognizer," *IEEE Trans. Inf. Technol. Biomed.*, vol. 14, no. 5, pp. 1166–1172, Sep. 2010.

[18] T. Huynh and B. Schiele, "Analyzing features for activity recognition," in *Proc. Joint Conf. Smart Objects Ambient Intell.: Innovative Context-Aware Ser.: Usages Technol.*, Grenoble, France, 2005, pp. 159–163.

[19] J. Lester, T. Choudhury, N. Kern, G. Borriello, and B. Hannaford, "A hybrid discriminative/generative approach for modeling human activities," in *Proc. Int. Joint Conf. Artif. Intell.*, Edinburgh, U.K., 2005, pp. 766–772.

[20] M. Zhang and A. A. Sawchuk, "Motion primitive-based human activity recognition using a bag-of-features approach," in *Proc. 2nd ACM SIGHIT Int. Health Inf. Symp.*, Miami, FL, USA, Jan. 2012, pp. 631–640.

[21] J. Blackburn and E. Ribeiro, "Human motion recognition using isomap and dynamic time warping," in *Proc. Workshop Human Motion: Understanding, Model., Capture Animat.*, Rio de Janeiro, Brazil, Oct. 2007, pp. 285–298.

[22] M. A. Turk and A. P. Pentland, "Face recognition using eigenfaces," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Honolulu, HI, USA, Jun. 1991, pp. 586–591.

[23] P. Belhumeur, J. Hespanha, and D. Kriegman, "Eigenfaces vs. fisherfaces: Recognition using class specific linear projection," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 19, no. 7, pp. 711–720, Jul. 1997.

[24] E. Bingham and H. Mannila, "Random projection in dimensionality reduction: Applications to image and text data," in *Proc. Int. Conf. Knowl. Discovery Data Mining*, San Francisco, CA, USA, 2001, pp. 245–250.

[25] E. Candes, "Compressive sampling," in *Proc. Int. Congr. Math.*, Madrid, Spain, Aug. 2006, pp. 1433–1452.

[26] E. Candes, J. Romberg, and T. Tao, "Robust uncertainty principles: Exact signal reconstruction from highly incomplete frequency information," *IEEE Trans. Inf. Theory*, vol. 52, no. 2, pp. 489–509, Feb. 2006.

[27] D. Donoho, "Compressed sensing," *IEEE Trans. Inf. Theory*, vol. 52, no. 4, pp. 1289–1306, Apr. 2006.

[28] S. S. Chen, D. L. Donoho, and M. A. Saunders, "Atomic decomposition by basis pursuit," *SIAM Rev.*, vol. 43, pp. 129–159, Jan. 2001.

Mi Zhang (S'12) received the B.S. degree in electrical engineering from Peking University, Beijing, China, in 2006, and the M.S. degree in electrical engineering and the M.S. degree in computer science from the University of Southern California, Los Angeles, CA, USA, in 2010. He is currently working toward the Ph.D. degree in the Ming Hsieh Department of Electrical Engineering, University of Southern California, Los Angeles, CA.

His research interests include ubiquitous and mobile computing, embedded sensing systems, machine intelligence, smart environment, with a particular focus on personal healthcare and medical applications.

Alexander A. Sawchuk (LF'11) received the B.S. degree in electrical engineering from the Massachusetts Institute of Technology, Cambridge, MA, USA, and the M.S. and Ph.D. degrees in electrical engineering from Stanford University, Stanford, CA, USA.

He is a Leonard Silverman Chair Professor in the Ming Hsieh Department of Electrical Engineering, University of Southern California (USC), Los Angeles, CA. He has served in the past as the Director of the Signal and Image Processing Institute, and as the Deputy Director of the Integrated Media Systems Center. His research interests include the use of optoelectronic devices and systems for parallel optical computing, interconnection, network and data storage applications. Other interests include image processing, immersive media technology, stereo and panoramic video and displays; wireless health, body area networks, multimodal sensing, and pattern recognition. He is the holder of two U.S. Patents and is the author or coauthor of more than 280 technical publications in these fields.

Dr. Sawchuk is a Fellow of the Optical Society of America (OSA) and the International Society for Optical Engineering (SPIE). He received the USC Associates Award for Excellence in Teaching, the OSA Distinguished Service Award, and the USC Mellon Academic Mentoring Support Program Certificate of Recognition. He also received the Halliburton Award for Exceptional Service, the Lockheed Senior Research Award, and the Outstanding Teaching Award from the USC Viterbi School of Engineering. He has served often in various offices and on technical committees for IEEE, OSA, and SPIE. He was an Associate Editor for the OSA Journal *Optics Letters*, and was the Editor of the Information Processing Division of *Applied Optics*.