

# Robust Human Activity and Sensor Location Corecognition via Sparse Signal Representation

Wenyao Xu\*, *Student Member, IEEE*, Mi Zhang, *Student Member, IEEE*, Alexander A. Sawchuk, *Fellow, IEEE*, and Majid Sarrafzadeh, *Fellow, IEEE*

**Abstract**—Human activity recognition with wearable body sensors receives lots of attentions in both research and industrial communities due to the significant role in ubiquitous and mobile health monitoring. One of the most concerned issues related to this wearable technology is that the sensor signals significantly depends on where the sensors are worn on the human body. Existing research work either extracts location information from the activity signals or takes advantage of the sensor location information as *a priori* information to achieve better activity recognition performance. In this paper, we present a sparse signal-based approach to corecognize human activity and sensor location in a single framework. Therefore, the wearable sensor is not necessarily constrained to fixed body position and the deployment is much easier although the recognition difficulty becomes much more challenging. To validate the effectiveness of our approach, we run a pilot study in the lab, which includes 14 human activities and seven on-body locations to recognize. The experimental results show that our approach achieves an 87.72% classification accuracy (the mean of precision and recall), which outperforms classical classification methods.

**Index Terms**—Body sensor networks, human activity recognition, inertial sensor, sensor localization, sparse representation.

## I. INTRODUCTION

DAILY activities and behaviors is a key indicator of human health status. In recent years, a lot of clinical studies prove that human activity and behaviors is highly related to medical diseases, such as obesity [1], autism [2], and cerebral palsy [3]. Therefore, it is important to enable accurate human activity measurements, and proven data can be used as evidence for medical diagnosis and treatment.

Currently, there are two main methods to monitor human activities. One is to deploy the remote sensors (such as camera, radar, and infrared) in infrastructure such as parking lots, office buildings, and airports. Park and Trivedi [4] presented a camera-based tracking system for human movement analysis and privacy protection. Kim and Ling [5] investigated the feasibility of classifying different human activities based on micro-Doppler

signatures. Han and Bhanu [6] developed a hierarchical scheme to automatically detect human movements via infrared sensors. However, one commonly key drawback is that the sensing scope of these remote sensors is immobile and limited due to fixed deployment position. Information may be lost if people leave the field of view of the sensors. Moreover, because of the privacy issue, users will feel uncomfortable when their daily activities are under monitor.

The other method is to attach the inertial sensors (accelerometer, gyroscope, etc.) on the human body to sensor the motor of body parts. Bouten *et al.* [7] described the development of a triaxial accelerometer and a portable data processing unit for the assessment of daily physical activity. Najafi *et al.* [8] introduced a gyroscope-based ambulatory system to detect body postures in elderly persons. With advances in microelectromechanical system technologies, inertial sensors have become popular in recent years for human activity sensing and tracking since they are low cost, miniature, and easy to be integrated into personal carry-on devices such as smart phones, watches, and apparel (e.g., shoes, clothes, and hats). Fig. 1 illustrates a number of examples of on-body sensing devices integrated with inertial sensors. We can see that different devices will be deployed on different locations on the human body accordingly. Therefore, the activity signals captured by inertial sensors are highly dependent on both the human activity and the location where the sensors are worn. In other words, it is highly possible that the signals may be totally different when a person performs the same activity with sensors on different locations. Therefore, any misplacement or displacement might lose the sensing accuracy and make the analysis completely incorrect [9].

Based on this observation, it is difficult to understand the inertial sensor signal without any *priori* information. In fact, researchers have been developing techniques to either extract location information from the captured activity signals or take advantage of the sensor location information as *a priori* information to achieve better activity recognition performance. For example, Vahdatpour *et al.* [10] developed a support vector machine (SVM)-based approach to identify the sensor location on the human body when people walk. Long *et al.* [11] customized the activity recognition algorithm to specific sensor locations to improve the performance of the recognizer. Among all these existing techniques, the common point is that they treat sensor localization and activity recognition as two separate problems assuming the other part information is known. Unfortunately, either of them will be available in practical applications. In this paper, we argue that the sensor location information and the activity signals are intertwined and can be solved as one problem.

Manuscript received February 18, 2012; revised April 27, 2012; accepted May 23, 2012. Date of publication August 3, 2012; date of current version October 16, 2012. Asterisk indicates corresponding author.

\*W. Xu is with the University of California at Los Angeles, Los Angeles, CA 90095 USA (e-mail: wxu@ee.ucla.edu).

M. Zhang and A. A. Sawchuk are with the University of Southern California, Los Angeles, CA 90089 USA.

M. Sarrafzadeh is with the Wireless Health Institute, University of California at Los Angeles, Los Angeles, CA 90095 USA.

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/TBME.2012.2211355



Fig. 1. Examples of on-body inertial sensing devices for human activity monitoring and recognition: 1) Nike+; 2) BodyMedia; 3) Healthset; 4) Basis band; 5) Fitbit.

We proposed the solution to corecognize human activity and sensor location in a single framework, which is not necessary to have any priori information about sensor and human activity. Specifically, our framework uses sparse signal theory, which enables us to reconstruct the signal with limited or incomplete samples if the signal has *sparsity* in some transformation domain [12]. Here, we prove that human activity signals captured by wearable inertial sensors are indeed sparse, and then take advantage of this sparsity information to classify activity signals and recognize where the sensor is located on the human body simultaneously. Based on our experiments, our method can recognize 14 activities and seven on-body locations with 87.72% recognition accuracy on average.

The remainder of this paper is organized as follows. In Section II, we briefly introduce the theory of sparse signals and review some existing work on human activity recognition based on the theory. Then, we describe our human activity and sensor location corecognition framework in Section III. The experimental results and analysis are presented in Section IV. Finally, we outline the future work and conclude the paper in Section V.

## II. PRELIMINARY AND RELATED WORK

### A. Sparse Signals and $\ell_1$ Minimization

Sparse signal processing (or compressed sensing) [12] is a ground-breaking signal processing procedure developed in recent years. It has been widely applied in many research domains, such as communications, medical image processing, and computer graphics, due to its capability of accurate signal reconstruction with lower sampling rate claimed by Nyquist–Shannon sampling theorem [13].

Suppose that  $x \in \mathbb{R}^n$  is a vector of unknown variables,  $y \in \mathbb{R}^m$  is the available measurements, and  $A \in \mathbb{R}^{m \times n}$  is the data

matrix to describe the relation between  $x$  and  $y$ . Then, we have

$$y = Ax. \quad (1)$$

In many real-world applications, the number of unknowns  $n$  is more than the number of measurements,  $m$ , i.e.,  $n > m$ . In such cases, (1) represents an underdetermined system, and  $x$  cannot be uniquely reconstructed from the data matrix  $A$  and the measurements  $y$ . However, in situations where  $x$  is sparse enough, we can reconstruct  $x$  with the  $\ell_0$  sparsity formulation from

$$\begin{aligned} \min_{x \in \mathbb{R}^n} \|x\|_{\ell_0} \\ \text{s.t. } y = Ax. \end{aligned} \quad (2)$$

Equation (2) is a determined system and has a stable solution. However, (2) is intractable because it is an NP-hard problem [14]. The traditional heuristic to approximate the sparsity  $\ell_0$  is to use the minimal energy  $\ell_2$

$$\begin{aligned} \min_{x \in \mathbb{R}^n} \|x\|_{\ell_2} \\ \text{s.t. } y = Ax. \end{aligned} \quad (3)$$

It is well known that  $\ell_2$  is a least square formation and can be efficiently solved. However, the energy minimization  $\ell_2$  is not necessarily equivalent to the sparsity  $\ell_0$  in most cases. In 2006, the authors in [13] proved that the solution of (2) is highly the same as the solution with the  $\ell_1$  minimization

$$\begin{aligned} \min_{x \in \mathbb{R}^n} \|x\|_{\ell_1} \\ \text{s.t. } y = Ax. \end{aligned} \quad (4)$$

It has been proved that this  $\ell_1$  minimization can be formulated as a convex problem [12]. In such case, the optimization problem is well posed and can be solved in polynomial time.

### B. Sparse Signals for Pattern Recognition

One important application of sparse signals is pattern recognition and classification. In recent years, it has been applied successfully to many pattern recognition problems including face recognition [15], speech recognition [16], and iris recognition [17]. The formulation of the sparse signal-based classification strategy is relatively straightforward. Consider a pattern recognition problem with  $K$  different classes. Each class  $k$  has  $n_k$  training samples, each having  $m$  attributes. In total, there are  $n = \sum_{i=1}^K n_i$  training samples. We can collect all these training samples to form a matrix  $A$  with  $m$  rows and  $n$  columns as follows:

$$\begin{aligned} A &= [A_1, A_2, \dots, A_i, \dots, A_k] \\ &= [a_{11}, a_{12}, \dots, a_{1n_1}, a_{21}, a_{22}, \dots, a_{2n_2}, \dots, \\ &\quad \dots, a_{i1}, a_{i2}, \dots, a_{in_i}, \dots, a_{k1}, a_{k2}, \dots, a_{kn_k}] \end{aligned} \quad (5)$$

where  $a_{ij}$  is the  $j$ th training sample from class  $i$ .

Following (1), any given unknown input  $y \in \mathbb{R}^m$  can be represented as a linear span of training sample matrix  $A \in \mathbb{R}^{m \times n}$  as

$$y = x_1 a_{11} + x_2 a_{12} + \dots + x_n a_{kn_k} \quad (6)$$

where  $x_1, x_2, \dots, x_n$  are a sparse set of weights.



Fig. 2. Three important components of our sparse signal-based framework.

Under this formulation, the class membership of  $y$ , which is encoded as the sparsest solution of the underdetermined system given in (1), can be found by solving (4).

### C. Related Work

There are some existing research work on using sparse representation (SR) for human activity analysis. Yang *et al.* [18] used eight motion sensing nodes distributed on the human body to recognize 12 different human activities. Liu *et al.* [19] adopted a similar strategy to recognize human activities captured by camera videos. Compared to these existing studies, our work differs in the following aspects.

- 1) *Sensing technology*: Instead of using video cameras [19], we use inertial sensors (accelerometer and gyroscope) attached on the human body to collect human activity signals.
- 2) *Sensing strategy*: Rather than distributing multiple sensor nodes to cover the entire body as in [18], we use only one single sensor node on the body to recognize human activity. We believe this sensing strategy is less intrusive and can be applied to a much wider range of real-world applications.
- 3) *Sensor location*: The work in [18] requires that the sensor nodes are fixed in some specific locations. Any misplacement of the sensor nodes will make the recognition fail. In comparison, our method does not have this limitation and enables the corecognition of human activity and sensor location in one single step.

A preliminary version of this work has appeared [20]. Compared to previous work, we include a further study and discussion of the algorithm performance in this paper. Specifically, there are two new contributions. First, we investigated algorithm *robustness* on feature dimension reduction. In this paper, we evaluate the algorithm robustness from two aspects: classification *accuracy* and classification *stability*. The detailed analysis of this part is addressed in Section IV-D. Second, we compared our algorithm with two most used classification methods, nearest neighbor (NN) and nearest subspace (NS), to justify its advantage. The comparison study includes three aspects: classification accuracy, classification stability, and robustness on feature dimension reduction. The experimental results proved the advantage of our method and discussed the applicable conditions (see Section IV-F). Note that we use the same dataset on evaluation for the sake of consistency with previous work.

## III. OUR FRAMEWORK

In this section, we present our proposed framework for corecognizing human activity and on-body sensor location. As shown in Fig. 2, our framework consists of three important components: feature extraction, SR via  $\ell_1$  minimization, and Bayesian sparse

signal-based classification. We will describe the details of all these components in this section.

### A. Feature Extraction

There are many existing studies focusing on exploring the best features that can be extracted from human activity signals. Following previous work [21], we extract statistical features in motion signals, such as mean, standard deviation, and nonnegative entropy, to represent human activities. These parameters index a family of probability distribution and describe the characteristics of a population. In this paper, we use 12 statistical features listed in Table I because these features have been proven to be useful in classifying human activities and other related pattern recognition problems by existing studies [22]. All these features are extracted from both accelerometer and gyroscope signals. In total, 64 features for each segment are extracted and denoted as  $x \in \mathbb{R}^{64}$ . Also, note that every feature in the vector is independent of others.

### B. SR via $\ell_1$ Minimization

We follow the formulation described in Section II to construct the data matrix  $A$ . Specifically, we collect  $n$  samples from activity  $i$  and sensor location  $j$ . For each sample, we extract features described in the previous section to form a feature vector  $a$ . Then, a feature matrix  $A_{ij}$  can be constructed as

$$A_{ij} = [a_1, a_2, \dots, a_n]. \quad (7)$$

In this way, we build the data matrix  $A$  covering all  $K$  activities and  $L$  locations as

$$A = [A_{11}, A_{12}, \dots, A_{KL}]. \quad (8)$$

As described in Section II, any given test sample  $y$  from unknown activity and location can be represented in terms of the matrix  $A$  as

$$y = A_{11}x_{11} + A_{12}x_{12} + \dots + A_{KL}x_{kl} \quad (9)$$

where  $x = [x_{11}, x_{12}, \dots, x_{kl}]$  is the SR of  $y$  with respect to matrix  $A$ , and the coefficient  $x_{ij}$  is referred as the feature index for feature matrix  $A_{ij}$ . In such case,  $x$  can be found via the  $\ell_1$  minimization formulated in (4).

### C. Bayesian Sparse Signal-Based Classification

Given the SR  $x$  of the test sample  $y$ , we can identify its activity class membership  $i$  and location class membership  $j$  altogether by computing the residual values between  $y$  and each feature matrix  $A_{ij}$  defined as

$$\mathbf{residual}_{ij} = \|y - A_{ij}x_{ij}\|_2 \quad (10)$$

the lower the residual value is, the more similar  $y$  is to feature matrix  $A_{ij}$ . Therefore,  $y$  is classified as the activity class  $C$  and

TABLE I  
SELECTED FEATURES AND THEIR DEFINITIONS

No.	Statistical Feature	Definition
1	Arithmetic Mean	The DC component (average value) of the signal
2	Median	The median signal value
3	Stand Deviation	Measure of the distribution of the signal
4	Variance	The square of the standard deviation
5	Root Mean Square	The quadratic mean value of the signal
6	Mean Derivatives	The average value of the first order derivatives of the signal over the window
7	Skewness	The degree of asymmetry of the sensor signal distribution
8	Kurtosis	the degree of peakedness of the sensor signal distribution
9	Interquartile Range	Measure of the statistical dispersion, being equal to the difference between the 75th and the 25th percentiles of the signal over the window
10	Zero Crossing Rate	The total number of times the signal changes from positive to negative or vice versa
11	Mean Crossing Rate	The total number of times the signal change from below average to above average or vice versa
12	Pairwise Correlation	Correlation between two axes of each sensor and different sensors

sensor location class  $S$  that produces the smallest residual

$$\{C, S\} = \arg \min_{ij} \text{residual}_{ij}. \quad (11)$$

Let  $P(i, j|C, S)$  be the probability that the test sample  $y$  is classified as activity  $i$  and sensor location  $j$  when the true activity class is  $C$  and true sensor location class is  $S$ . Since the residual value is a measure of the similarity between  $y$  and the feature matrix  $A_{ij}$ , the lower the residual, the higher the probability that the classified activity class  $i$  and location class  $j$  will be the same as the true activity class  $C$  and true location class  $S$ . Therefore, we can model the probability  $P(i, j|C, S)$  using the residual values as

$$P(i, j|C, S) = 1 - \frac{\text{residual}_{ij}}{\sum_{i=1}^K \sum_{j=1}^L \text{residual}_{ij}}. \quad (12)$$

Based on the sum rule of probability theory, the probability of  $y$  classified as activity  $i$  when the true activity class is  $C$  can be derived by summing up the probability at each sensor location from

$$P(i|C) = 1 - \frac{\sum_{j=1}^L \text{residual}_{ij}}{\sum_{i=1}^K \sum_{j=1}^L \text{residual}_{ij}} \quad (13)$$

and the test sample  $y$  is classified as activity class  $C^*$  that has the highest probability

$$C^* = \arg \max_i P(i|C). \quad (14)$$

Similarly, the probability of  $y$  classified as location  $j$  when the true location class is  $S$  is calculated by summing up the probability over all activity classes from

$$P(j|S) = 1 - \frac{\sum_{i=1}^K \text{residual}_{ij}}{\sum_{i=1}^K \sum_{j=1}^L \text{residual}_{ij}} \quad (15)$$

and the test sample  $y$  is classified as sensor location class  $S^*$  that has the highest probability

$$S^* = \arg \max_j P(j|S). \quad (16)$$

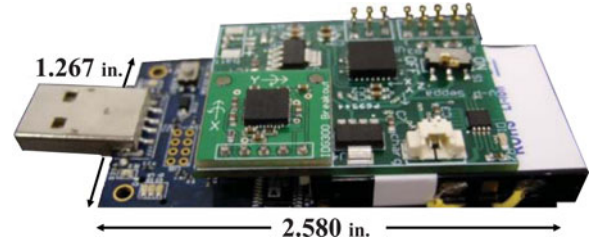


Fig. 3. TelosB mote for human motion sensing.

## IV. EXPERIMENTS AND EVALUATION

### A. Experimental Setup and Dataset

We ran a pilot study in a laboratory environment to evaluate the performance of our proposed approach. The dataset is collected using an customized human motion sensing node, called a TelosB mote [23]. As shown in Fig. 3, the size of the TelosB mote is 1.267 in by 2.580 in, and it includes a triaxial accelerometer and a biaxial gyroscope (i.e., *five* data channels). For each channel, the signal sampling rate is 50 Hz, and all the samples are stored in an on-board SD card.

We collected the data from three male subjects whose ages are 25, 28, and 33, respectively. Each subject performed 14 different activities including: 1) stand to sit; 2) sit to stand; 3) sit to lie; 4) lie to sit; 5) bend to grasp; 6) rising from bending; 7) kneeling right; 8) rising from kneeling; 9) look back; 10) return from look back; 11) turn clockwise; 12) step forward; 13) step backward; and 14) jumping. Meanwhile, the subjects were asked to wear the sensing device at seven different locations during their performance. These locations are a) waist; b) right wrist; c) left wrist; d) right arm; e) left thigh; f) right ankle; and g) left ankle. Therefore, we have 98 combinations of activity and sensor location in total. For each location, each subject performed each activity for ten trials.

After data collection, we extract features listed in Table I from each data unit. Our sensor node has *five* data channels, and we can extract 64 features from each data unit.

### B. Sparsity of Human Activity

Based on the discussion in Section II-A, SR can perform accurate recognition and classification based on one important

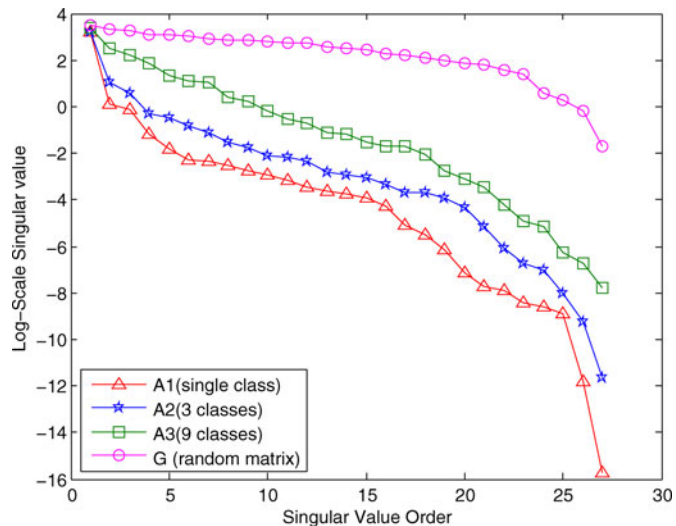


Fig. 4. Log-scale singular values of the sample matrix  $A_1$ ,  $A_2$ , and  $A_3$ . We also use Gaussian random matrix  $G$  for comparison.

prerequisite: *the representation  $x$  of  $y$  should be a sparse vector in the space spanned by the training samples  $A$ .* Unfortunately, few works prove the sparsity of their problems before using this principle, either theoretically or empirically. For the sake of avoiding blind decisions, we did the preliminary experiments to investigate the sparsity of human activities.

Without the loss of generality, we randomly selected 30 samples from a single activity, and each sample has 30 randomly selected features. In this way, we can form a sample matrix  $A_1 \in \mathbb{R}^{30 \times 30}$ . We also built  $A_2 \in \mathbb{R}^{30 \times 30}$  with three human activities and  $A_3 \in \mathbb{R}^{30 \times 30}$  with nine activities. Note that in all these sample matrixes, column space consists of samples, and the row space is based on the features. Similar to [24], we generated a Gaussian random matrix  $G \in \mathbb{R}^{30 \times 30}$  and performed singular value decomposition (SVD) [25] on  $A_1$ ,  $A_2$ ,  $A_3$ , and  $G$ , respectively, to evaluate the sparsity of human activity. All their singular value profiles are illustrated in Fig. 4. It indicates that compared to  $G$ ,  $A_1$ ,  $A_2$ , and  $A_3$  are low rank since their SVD profiles have significant downtrend compared to  $G$ . Knowing that all features are statistically independent (see Section III-A), low rank should be caused by column space, which means that the sample space is overcomplete. This is an experimental indication that the training samples  $A$  are a sparse set. Specifically, comparing  $A_1$  with  $A_2$  and  $A_3$ , we can see that the sample space from the same activity class (i.e.,  $A_1$ ) has a higher degree of sparsity.

### C. Classification Performance Evaluation

For this part, we evaluate the classification performance of our framework. Our evaluation is based on three metrics: 1) the classification accuracy of corecognition of activity and sensor location based on (11); 2) the classification accuracy of activity based on (14); and (3) the classification accuracy of sensor location based on (16). For evaluation, we adopt a tenfold cross-validation strategy. Specifically, we divide the whole dataset into ten sets. At one time, five sets are used to build the data matrix  $A$  and the remaining five sets for testing. The whole procedure iterates ten times.

There are two important aspects in algorithm evaluation. One is the classification *accuracy*. Given large number of testing inquiries, the algorithm should offer the correct responses with high probability. The other aspect is the classification *stability*. If the training and testing datasets are changed, the recognition result should vary slightly from the average rate.

Table II shows the evaluation results in terms of the aforementioned three metrics. As shown, metric 1 achieves an 87.42% precision value and an 87.93% recall value. For metrics 2 and 3, it is interesting to see that with Bayesian fusion, the classification accuracy is improved. Specifically, for activity recognition, the precision and recall reach 88.79% and 89.21%. For location recognition, both the precision and the recall are higher than 96%. For *stability*, we observe that the standard deviation of metric 1 is as low as 1.26% (the mean of precision and recall variation). After Bayesian fusion, both metrics 2 and 3 have better robustness. As shown in Table II, the variation of metric 2 is 1.25% and 1.02% for precision and recall variation. For metric 3, the variation is always less than 0.5% in general.

To take a closer look at the classification results, Tables III and IV show two confusion tables with respect to activity classification (metric 2) and sensor location classification (metric 3), respectively. In Table III, we notice that activity 7) *kneeling right* and activity 8) *rising from kneeling* are most often confused. Although these two activities are a pair of complementary processes and visibly different from each other in the *time domain*, our algorithm describes the human activity signal in a statistical way and removes the temporal information in the data. Therefore, complementary processes should share many features in the *space domain*. As for sensor location classification, as illustrated in Table IV, most precision and recall values are more than 98%. However, location e) *Left Thigh* and location f) *Right Ankle* are confused with each other the most. Specifically, the corresponding accuracy is around 92%. It indicates that the selected features described in Section III cannot reliably distinguish the two cases. We could consider this issue to enhance the algorithm performance in future work.

### D. Robustness on Feature Dimension Reduction

In this experiment, we investigate the impact of feature dimension on performance of our framework. As mentioned previously, in total we extract 64 features. We start from 60 features and reduce the number of features by the decrement 5. Fig. 5 shows the averaged misclassification rates for each feature dimension in terms of the three classification metrics. We can see that the misclassification rate of corecognition increases from 13% to 39%, while the feature dimension decreasing from 64 to 30. Furthermore, it indicates that 55 is the significant turning point, where the misclassification rates increase greatly (more than 10%) when the feature dimension is less than 55 (from 55 to 50).

In addition to the classification accuracy (see Fig. 5), we also examined the classification stability of our framework with feature dimension. For simplicity of the presentation, we investigated the performance on corecognition of human activity and sensor location (metric 1). As shown in Fig. 6 (the blue error bar), we observe that the variation of the misclassification

TABLE II  
CLASSIFICATION PERFORMANCE EVALUATED BY THREE METRICS

	Activity&Location (%) metric (1)	Activity (%) metric (2)	Location (%) metric (3)
Precision	87.42 ± 1.43	88.79 ± 1.25	96.02 ± 0.43
Recall	87.93 ± 1.10	89.21 ± 1.02	96.24 ± 0.38

TABLE III  
CONFUSION TABLE OF RECOGNITION ON 14 HUMAN ACTIVITIES

	1	2	3	4	5	6	7	8	9	10	11	12	13	14	Total	Recall
1	79	1	0	1	0	0	0	0	0	0	0	1	0	2	84	94%
2	3	79	0	0	0	0	0	0	0	2	0	0	0	0	84	94%
3	1	1	74	8	0	0	0	0	0	0	0	0	0	0	84	88%
4	2	2	3	72	0	1	2	0	0	0	2	0	0	0	84	86%
5	0	0	0	1	78	0	0	0	0	1	0	0	1	2	84	93%
6	0	0	0	0	0	78	0	0	1	1	2	1	0	1	84	93%
7	0	0	0	0	0	0	72	8	0	0	2	1	1	0	84	86%
8	0	0	0	0	1	0	8	67	0	0	1	4	2	1	84	80%
9	0	0	0	0	1	0	0	0	78	2	3	0	0	0	84	93%
10	1	0	0	0	0	2	0	0	4	76	1	0	0	0	84	90%
11	1	0	0	0	0	0	2	0	1	2	71	2	5	0	84	85%
12	0	0	0	2	0	0	4	0	0	0	1	74	3	0	84	88%
13	0	0	0	0	0	0	2	4	1	2	0	9	66	0	84	79%
14	0	0	0	0	0	0	0	0	0	0	0	0	0	84	84	100%
<b>Total</b>	87	83	77	84	80	81	100	79	85	86	83	92	78	90		
<b>Precision</b>	91%	95%	96%	86%	98%	96%	72%	85%	92%	88%	86%	80%	85%	93%		

TABLE IV  
CONFUSION TABLE OF RECOGNITION ON SEVEN ON-BODY SENSOR LOCATIONS

	a	b	c	d	e	f	g	Total	Recall
a	166	1	0	0	1	0	0	168	99%
b	0	164	2	1	1	1	0	168	97%
c	2	1	158	0	4	1	2	168	94%
d	0	0	1	163	3	1	0	168	97%
e	4	0	0	0	154	10	0	168	92%
f	2	1	1	0	5	157	2	168	93%
g	0	0	0	0	0	0	168	168	100%
<b>Total</b>	174	166	162	164	168	170	172		
<b>Precision</b>	95%	99%	98%	99%	92%	92%	98%		

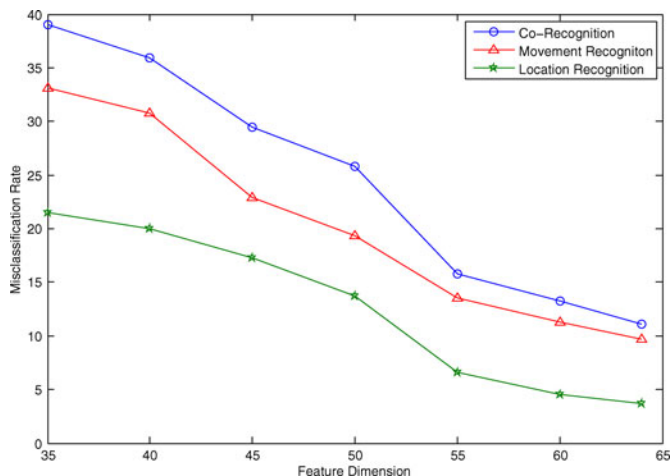


Fig. 5. Impact of feature dimension on classification accuracy.

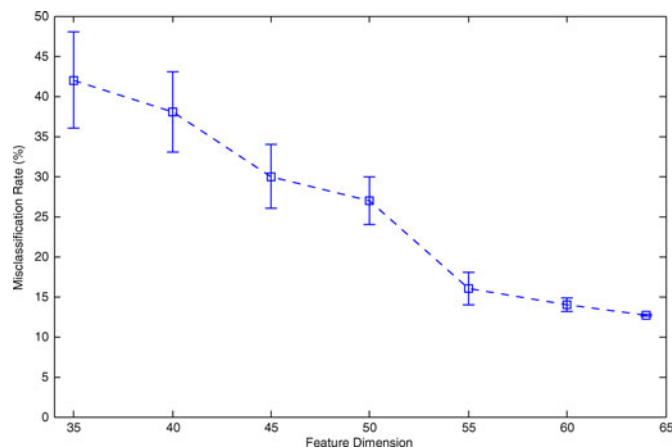


Fig. 6. Impact of feature selection on classification stability of metric 1.

rate also increases with a decrease of the number of features. This implies that the feature number has significant impact not only on the classification accuracy, but also on the classification stability. Specifically, the variation increases to 5% when the feature number drops to 35.

More specifically, to examine the robustness of our framework in terms of different features, we randomly selected 55 features out of 64 as an example and evaluated the framework for ten times. Table V shows the results. As shown, it is interesting to see that our compressed sensing-based framework is not sensitive

TABLE V  
IMPACT OF FEATURES ON CLASSIFICATION PERFORMANCE

	Activity&Location (%)	Activity (%)	Location (%)
Precision	83.14 ± 1.41	86.01 ± 1.15	92.73 ± 0.53
Recall	83.78 ± 1.27	87.15 ± 0.94	93.24 ± 0.47

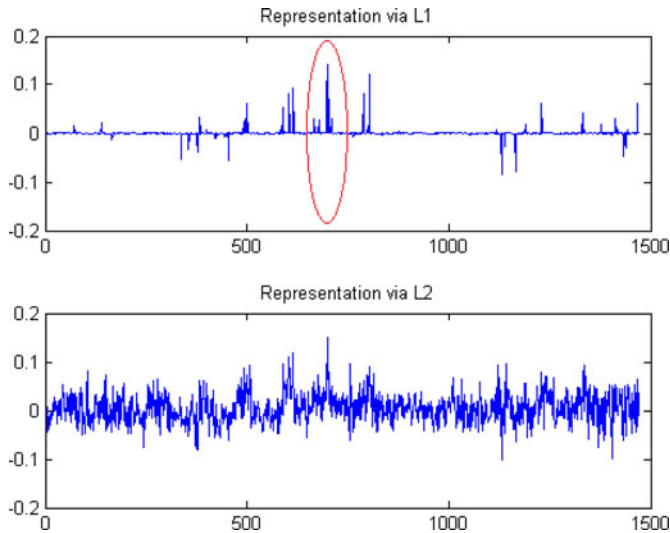


Fig. 7. Solutions of  $\ell_1$  and  $\ell_2$  optimization strategies.

to the specific features selected, with the standard deviation of the performance less than 1.41% for all cases.

### E. Comparison Between $\ell_1$ and $\ell_2$

As stated in Section II,  $\ell_1$  is a better heuristic for sparsity optimization than  $\ell_2$ . In this section, we empirically validate this point and compare the classification performance between  $\ell_1$  and  $\ell_2$  optimization strategies. As an example, Fig. 7 shows the solutions from both  $\ell_1$  and  $\ell_2$  optimization with one test sample from activity 7 (kneeling right) at location  $d$  (right arm). As illustrated, the solution from  $\ell_1$  is quite sparse and has few significantly large components. Moreover, the maximal spike marked by the red circle is associated with the training samples belonging to the same activity class and sensor location class. In comparison, the solution from  $\ell_2$  has few significant components and many small components. The spikes are dense and distributed over all activity and sensor location classes (see Fig. 7).

For clarity of the presentation, we augment this result in the residual computation. Fig. 8 illustrates the corresponding residual values between the test sample and all 98 classes defined by (10) for both  $\ell_1$  and  $\ell_2$ . As shown in the figure, the class membership of the test sample can be easily found by the minimal residual (pointed by the red arrow) for the  $\ell_1$  optimization strategy. For  $\ell_2$ , although the minimal residual also corresponds to the true class, the difference between the minimal residual and the residual values of other classes is not significant in this example.

Also, we compare the classification performance between  $\ell_1$  and  $\ell_2$ . Table VI shows the results in terms of the recognition rates. As shown,  $\ell_1$  outperforms  $\ell_2$  across all three metrics consistently in terms of both recognition accuracy and stability. It

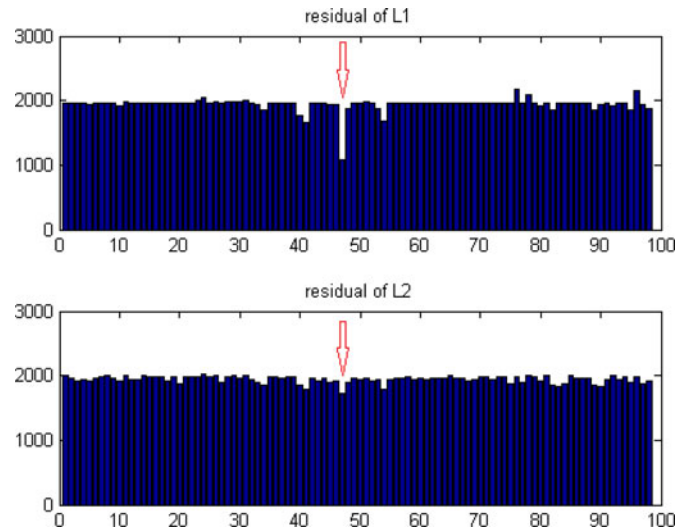


Fig. 8. Residuals of 98 classes of  $\ell_1$  and  $\ell_2$  optimization strategies.

TABLE VI  
CLASSIFICATION PERFORMANCE COMPARISON OF  $\ell_1$  AND  $\ell_2$

	Activity&Location metric (1)		Activity metric (2)		Location metric (3)	
	mean	std	mean	std	mean	std
$\ell_1$ (%)	87.72	1.26	89.00	1.13	96.13	0.41
$\ell_2$ (%)	72.65	5.46	80.94	4.28	85.32	1.31
$\frac{\ell_1 - \ell_2}{\ell_1}$	20.75%	3.17X	9.95%	2.78X	11.25%	2.20X

is worth emphasizing that the enhancement from  $\ell_1$  compared to  $\ell_2$  has stronger scalability: the larger the scale, the greater is the benefit. Based on the indication in Fig. 7, it is not surprised that  $\ell_1$  outperforms  $\ell_2$  overwhelmingly in terms of both accuracy and stability. Specifically, the corecognition classification accuracy could be improved by 20.75% with  $\ell_1$  optimization. Correspondingly, the gain of stability from  $\ell_1$  optimization is 3.17X on average.

### F. SR-Based Classifier Versus Two Classical Methods

As a last experiment, we show the advantage of our method with respect to different feature dimensions compared to two classical classification methods: NN-based classifier [26] and NS-based classifier [27]. For simplicity of the presentation, we only show the results on classification accuracy (the mean of precision and recall) here. Fig. 9 illustrates the average classification accuracy rates as a function of feature dimension. Each curve represents one classification method, respectively. Similar to Section IV-D, features at each dimension are sequentially selected, and the performance is based on the average.

As shown in the figure, all methods have better performance with feature dimension scaling-up. We can see that the performance of NN and NS is close, which follow the similar trend, and the SR-based classifier is different from them. More specifically, when the feature dimension is less than 40, NN and NS are better than SR; when the feature dimension is more than 40, the SR-based classifier outperforms the other two methods. This observation indicates that using only 40 features is not sufficient to recover the human activity signal via  $\ell_1$  minimization.

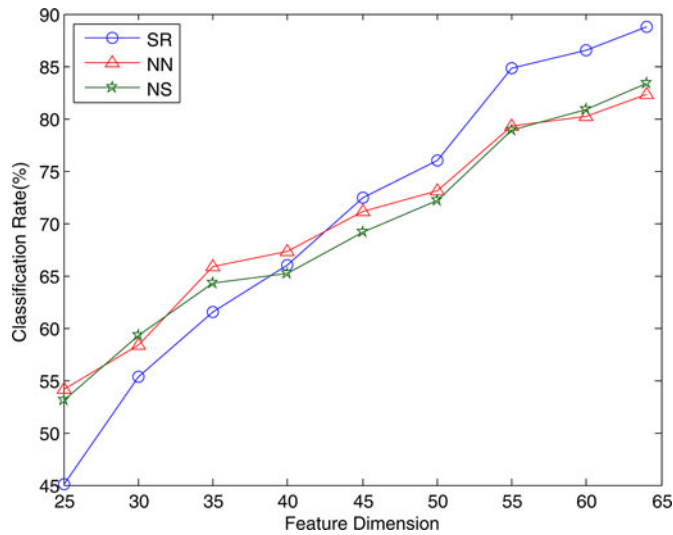


Fig. 9. Classification performance comparison between SR, NN, and NS.

However, when the feature dimension is equal or more than 45, the SR-based method can achieve a steady performance and beats the other two classical classification methods, achieving a maximum recognition rate of 87.72%.

## V. CONCLUSION

Inspired by the sparsity of human activity signals, we adopted and described a sparse signal representation technique to corecognize human activity and sensor location in wearable sensor networks. The experimental results showed that our method can achieve greater than 87% recognition accuracy with 14 different activities and seven on-body locations. We also showed that using the  $\ell_1$  norm is better than the  $\ell_2$  norm in terms of both accuracy and stability. Moreover, we compared an SR-based framework to two classical methods, NN-based method and NS-based method, respectively. The results showed that SR outperforms both of them when feature dimension is more than a threshold. Considering the promising results in the pilot study, we intend to run additional experiments with a larger scale group and evaluate more activities and sensor location in future work.

## ACKNOWLEDGMENT

The authors would like to thank Dr. R. Jafari, Director of Embedded Systems and Signal Processing Lab, University of Texas at Dallas, and Dr. H. Ghasemzadeh at Wireless Health Institute, University of California at Los Angeles, for the experimental equipments, dataset, and discussion.

## REFERENCES

- [1] M. Yoshioka, M. Ayabe, T. Yahiro, H. Higuchi, Y. Higaki, J. Amand, H. Miyazaki, Y. Yoshitake, M. Shindo, and H. Tanaka, "Long-period accelerometer monitoring shows the role of physical activity in overweight and obesity," *Int. J. Obesity*, vol. 29, pp. 502–508, Jun. 2005.
- [2] K. Liu, E. Brien, J. Guralnik, M. Criqui, G. Martin, P. Greenland, and M. McDermott, "Measuring physical activity in peripheral arterial disease: A comparison of two physical activity questionnaires with an accelerometer," *Angiology*, vol. 51, pp. 91–100, Dec. 2000.
- [3] S. Chen, C. Cunningham, B. Bennett, and J. Lach, "Enabling longitudinal assessment of ankle foot orthosis for children with cerebral palsy," in *Proc. ACM Int. Conf. Wireless Health*, San Deigo, CA, Oct. 2011, pp. 1–10.
- [4] S. Park and M. Trivedi, "A track-based human movement analysis and privacy protection system adaptive to environmental contexts," in *Proc. IEEE Conf. Adv. Video Signal Based Surveillance*, Los Alamitos, CA, Sep. 2005, pp. 171–176.
- [5] Y. Kim and H. Ling, "Human activity classification based on micro-Doppler signatures using a support vector machine," *IEEE Trans. Geosci. Remote Sens.*, vol. 47, no. 5, pp. 1328–1337, Dec. 2009.
- [6] J. Han and B. Bhanu, "Fusion of color and infrared video for moving human detection," *Pattern Recognit.*, vol. 40, pp. 1771–1784, May 2007.
- [7] C. Bouten, K. Koekkoek, M. Verduim, R. Kodde, and J. Janssen, "A triaxial accelerometer and portable data processing unit for the assessment of daily physical activity," *IEEE Trans. Biomed. Eng.*, vol. 44, no. 3, pp. 136–147, Mar. 1997.
- [8] B. Najafi, K. Aminian, A. Paraschiv, F. Loew, C. Bula, and P. Robert, "Ambulatory system for human motion analysis using a kinematic sensor: Monitoring of daily physical activity in the elderly," *IEEE Trans. Biomed. Eng.*, vol. 50, no. 6, pp. 711–723, Jun. 2003.
- [9] N. Amini, M. Sarrafzadeh, A. Vahdatpour, and W. Xu, "Accelerometer-based on-body sensor localization for health and medical monitoring applications," *Pervas. Mobile Comput.*, vol. 7, pp. 746–760, Aug. 2011.
- [10] A. Vahdatpour, N. Amini, and M. Sarrafzadeh, "On-body device localization for health and medical monitoring applications," in *Proc. Int. Conf. Pervas. Comput. Commun.*, San Deigo, CA, Jan. 2011, pp. 37–44.
- [11] X. Long, B. Yin, and R. Aarts, "Single-accelerometer-based daily physical activity classification," in *Proc. IEEE Int. Conf. Eng. Med. Biol. Soc.*, Minneapolis, MN, Jun. 2009, pp. 6107–6110.
- [12] D. Donoho, "Compressed sensing," *IEEE Trans. Inf. Theory*, vol. 52, pp. 1289–1306, May 2006.
- [13] 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, Dec. 2006.
- [14] B. Natarajan, "Sparse approximate solutions to linear systems," *SIAM J. Comput.*, vol. 24, pp. 227–234, May 1995.
- [15] J. Wright, A. Yang, A. Ganesh, S. Sastry, and Y. Ma, "Robust face recognition via sparse representation," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 31, no. 2, pp. 746–760, Feb. 2009.
- [16] J. Gemmeke, T. Virtanen, and A. Hurmalainen, "Exemplar-based sparse representations for noise robust automatic speech recognition," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 19, no. 7, pp. 2067–2080, Apr. 2007.
- [17] 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. 1–14, Oct. 2011.
- [18] A. Yang, R. Jafari, S. Sastry, and R. Bajcsy, "Distributed recognition of human actions using wearable motion sensor networks," *J. Ambient Intell. Smart Environ.*, vol. 1, pp. 103–115, Jun. 2009.
- [19] C. Liu, Y. Yang, and Y. Chen, "Human action recognition using sparse representation," in *Proc. IEEE Int. Conf. Intell. Comput. Intell. Syst.*, Beijing, China, Aug. 2009, pp. 184–188.
- [20] 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 9th Int. Conf. Wearable Implantable Body Sens. Netw.*, London, U.K., May 2012, pp. 124–129.
- [21] 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 Informat. Symp.*, Miami, FL, Jan. 2012, pp. 631–640.
- [22] 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 Netw.*, Beijing, China, Nov. 2011, pp. 1–7.
- [23] H. Ghasemzadeh, V. Loseu, and R. Jafari, "Wearable coach for sport training: A quantitative model to evaluate wrist-rotation in golf," *J. Ambient Intell. Smart Environ.*, vol. 1, pp. 173–184, Apr. 2009.
- [24] Q. Shi, A. Eriksson, A. V. D. Hengel, and C. Shen, "Is face recognition really a compressive sensing problem?" in *Proc. IEEE Int. Conf. Comput. Vis. Pattern Recognit.*, Colorado Spring, Jun. 2011, pp. 553–560.
- [25] G. Golub and C. Reinsch, "Singular value decomposition and least squares solutions," *Numerische Math.*, vol. 14, pp. 403–420, Nov. 1970.
- [26] T. Cover and P. Hart, "Nearest neighbor pattern classification," *IEEE Trans. Inf. Theory*, vol. 13, no. 1, pp. 21–27, Jan. 1967.
- [27] 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.