

Co-Recognition of Human Activity and Sensor Location via Compressed Sensing in Wearable Body Sensor Networks

Wenyao Xu*, Mi Zhang[†], Alexander A. Sawchuk[†], and Majid Sarrafzadeh*

*Wireless Health Institute, University of California, Los Angeles, California, 90095, USA

[†]Signal and Image Processing Institute, Ming Hsieh Department of Electrical Engineering, University of Southern California, Los Angeles, California 90089 USA

Abstract—Human activity recognition using wearable body sensors is playing a significant role in ubiquitous and mobile computing. One of the issues related to this wearable technology is that the captured activity signals are highly dependent on the location where the sensors are worn on the human body. Existing research work either **extracts location information from certain activity signals** or takes advantage of the sensor location information as a priori to achieve better activity recognition performance. In this paper, we present a compressed sensing-based approach to co-recognize human activity and sensor location in a single framework. To validate the effectiveness of our approach, we did a pilot study for the task of recognizing 14 human activities and 7 on-body locations. On average, our approach achieves an 87.72% classification accuracy (the mean of precision and recall).

Keywords—Human Activity Analysis, Wearable Device, Sensor Localization, Compressed Sensing

I. INTRODUCTION

Understanding human activities and behaviors [1] is a key topic in ubiquitous computing research because of its wide range of applications in human computer interaction, security surveillance, sports engineering, and intelligent assistance for elderly people. Among all the technologies, camera is the mostly used sensing device to capture human activity. However, one of its key drawbacks is that cameras have to be deployed in infrastructure. In such case, people may get out of track if they are beyond the reach of the cameras.

In recent years, the advance of the MEMS technologies makes inertial sensors becoming popular for human activity sensing and tracking since they can be integrated into personal devices, such as smart phones, watches, and apparels. Figure 1 illustrates a number of examples of on-body sensing devices integrated with inertial sensors. It should be noticed that these sensors could be worn at any location on the human body. Therefore, the activity signals captured by the inertial sensors are highly dependent on the sensor location. In other words, it is highly possible that the signals may look totally different when a person performs the same activity but with sensors at different locations.

Based on this observation, researchers have developed techniques to either extract location information from the captured activity signals or take advantage of the sensor location information as a priori to achieve better activity

recognition performance. For example, the authors in [2] have developed a SVM-based approach to identify the sensor location on the human body when people walk. In [3], the authors customized the activity recognition algorithm to specific sensor locations to boost the performance of the recognizer. **Among these existing techniques, the common point is that they treat sensor localization and activity recognition as two separate problems.** However, we argue that the sensor location information and the activity signals are intertwined and could be solved as one single problem.

In this paper, we have developed solution to co-recognize human activity and sensor location in a single framework, which we believe is more efficient and powerful than utilizing two separate algorithms for the same. Specifically, our framework is built on top of the compressed sensing theory which reconstructs the signal with limited or incomplete samples if the signal has *sparsity* in some transformation domain [4]. We first prove that human activity signals captured by the wearable inertial sensors are indeed sparse. Then we take advantage of this sparsity information to classify activity signals and recognize where the sensor is located on the human body. Based on our experiment, our method can recognize 14 activities and 7 on-body locations with 87.72% recognition accuracy on average.

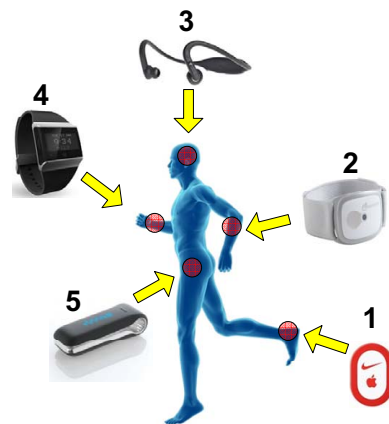


Figure 1. Examples of on-body inertial sensing devices for human activity monitoring and recognition: 1) Nike+; 2) BodyMedia; 3) Healthset; 4) Basis Band; 5) Fitbit.

The remainder of this paper is organized as follows. In Section II we briefly introduce the theory of compressed sensing and review some existing work on human activity recognition based on compressed sensing. Then, we describe our human activity and sensor location co-recognition 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. Compressed Sensing and ℓ_1 Minimization

Compressed sensing is a ground-breaking signal processing theorem developed in recent years. It has been widely applied in many research domains such as communication, image processing and computer graphics due to its capability of accurate signal reconstruction with lower sampling rate claimed by Nyquist-Shannon sampling theorem [5].

Suppose that $x \in \mathbb{R}^n$ is a vector of unknown variables, $y \in \mathbb{R}^m$ is the available measurements we have, 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 is more than the number of measurements ($n > m$). In such cases, Eq.1 represents an underdetermined system, and x can not 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 ℓ_0 sparsity formulation as follows:

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

Eq.2 is a determined system and has stable solution. However, Eq.2 is intractable because it is an NP-hard problem [6]. The traditional heuristic to approximate the sparsity ℓ_0 is to use the minimal energy ℓ_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 ℓ_2 is a least square formation and can be efficiently solved. However, the energy minimization ℓ_2 is not necessarily equivalent to the sparsity ℓ_0 in most cases. In 2006, authors in [5] proved that the solution of Eq.2 is highly probably the same with the ℓ_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 ℓ_1 minimization can be formulated as a convex optimization problem [4]. In such case, the optimization problem is well-posed and could be solved in polynomial time.

B. Compressed Sensing for Pattern Recognition

One important application of compressed sensing is pattern recognition and classification. In recent years, it has been applied successfully to many pattern recognition problems including face recognition, speech recognition, and iris recognition. The formulation of compressed sensing-based classification strategy is relatively straight forward. 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 Eq.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:

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

Under such formulation, the class membership of y , which is encoded as the sparsest solution of the underdetermined system given in Eq.1 can be resolved by solving Eq.4.

C. Related Work

There are some research work on using compressed sensing for human activity recognition. In [7], the authors used 8 motion sensing motes distributed on the human body to recognize 12 different human activities. In [8], the authors adopted a similar strategy to recognize human activities captured by camera videos. Compared to the existing studies, our work differ in the following aspects:

- 1) Sensing technology: Instead of using cameras, we use inertial sensors (accelerometer and gyroscope) attached on the human body to collect activity signals.
- 2) Sensing strategy: Rather than distributing multiple sensor nodes to cover the entire body like what [7] did, we only use one single sensor node on the body to recognize human activity. We believe this sensing strategy is less obtrusive and can be applied to a much wider range of real-world applications.
- 3) Sensor location: The work of [7] requires to fix the sensor nodes to 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 co-recognition of human activity and sensor location in one single step.

III. OUR FRAMEWORK

In this section, we present our proposed framework for co-recognizing human activity and on-body sensor location.



Figure 2. The three important components of our compressed sensing-based framework

As shown in Figure 2, our framework consists of three important components: feature extraction, sparse representation via ℓ_1 minimization, and Bayesian sparse representation-based classification. We will describe the details of all these components in this section.

A. Feature Extraction

There are many previous studies focusing on exploring the best features that can be extracted from human activity signals. Table I lists the features we consider in this work. We use these features because they have been proved to be useful in classifying human activities and some other related pattern recognition problems by existing studies [9].

Mean	Median	Standard Deviation
Variance	Root Mean Square	Mean Derivatives
Skewness	Kurtosis	Interquartile Range
Zero Crossing Rate	Mean Crossing Rate	Pairwise Correlation

Table I
FEATURES USED IN THIS WORK

B. Sparse Representation via ℓ_1 Minimization

We follow the formulation described in Section II to construct the data matrix A . Specifically, we collect n_{ij} samples from activity i and sensor location j . For each sample, we extract features described in the previous subsection 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 explained in Section II, for any given test sample y from unknown activity and location, it 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 sparse representation of y w.r.t. matrix A , and the coefficient x_{ij} is referred as feature index for feature matrix A_{ij} . In such case, x can be resolved via the ℓ_1 minimization formulated in Eq.4.

C. Bayesian Sparse Representation-Based Classification

Given the sparse representation 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:

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

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 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 of 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 is, 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 the 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:

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

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:

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

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

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

IV. EXPERIMENTS AND EVALUATION

A. Dataset

We run a pilot study in the laboratory environment to evaluate the performance of our proposed approach. The sensing device we used integrates a three axis accelerometer and a two axis gyroscope. Features listed in Table I are extracted from both of these sensors. In total, we have 64 features. We collected the data from 3 male subjects whose ages are 25, 28, and 33 respectively. Each subject performed 14 activities, each for 10 trials. The activities include: 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 7 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.

B. Sparsity of Human Activity

Based on the discussion in Section II-A, compressed sensing can perform accurate recognition and classification based on one important 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 problem before using compressed sensing theorem, 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 single activity, and each sample has 30 randomly selected features. In this way, we can form sample a sample matrix $A_1 \in \mathbb{R}^{30 \times 30}$. We also built $A_2 \in \mathbb{R}^{30 \times 30}$ with 3 human activities and $A_3 \in \mathbb{R}^{30 \times 30}$ with 9 activities. Note that in all these sample matrices, column space consists of samples, and row space is based on features. Similar to [10], we generated a Gaussian random matrix $G \in \mathbb{R}^{30 \times 30}$ and performed the singular value decomposition 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 Figure 3. 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 statistical features in Section III are independent, low-rank should be caused by column space, which means sample space is overcomplete. Therefore, we empirically proved that training samples A has sparsity. Specifically, comparing A_1 with A_2 and A_3 ,

we can see that sample space with the same activity class has the stronger sparsity.

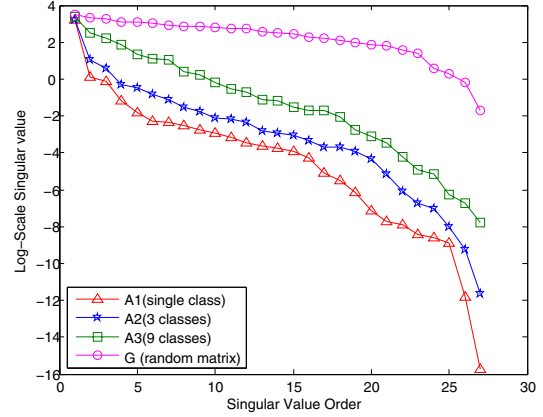


Figure 3. The log-scale singular values of the sample matrix A_1 , A_2 and A_3 . We also use Gaussian Random matrix G for comparison.

C. Classification Performance Evaluation

In this section, we evaluate the classification performance of our framework. Our evaluation is based on three metrics: (1) the classification accuracy of co-recognition of activity and sensor location based on Eq.11; (2) the classification accuracy of activity based on Eq.14; and (3) the classification accuracy of sensor location based on Eq.16. For evaluation, we adopt a 10-fold cross validation strategy. Specifically, we divide the whole dataset into 10 folds. At one time, 5 folds are used to build the data matrix A and the remaining 5 folds for testing.

Table II shows the evaluation results in terms of the above three metrics. As shown, metric (1) achieves an 87.42% precision and an 87.93% recall value. For metric (2) and (3), it is interesting to see that with Bayesian fusion, the classification performance gets 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%.

Table II
CLASSIFICATION PERFORMANCE EVALUATED BY THREE METRICS

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

To take a closer look at the classification results, Table 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* get confused to each other the most. Although these two activities are like a pair of inverse processes and visibly different

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%
Total	87	83	77	84	80	81	100	79	85	86	83	92	78	90		
Precision	91%	95%	96%	86%	98%	96%	72%	85%	92%	88%	86%	80%	85%	93%		

Table IV
CONFUSION TABLE OF RECOGNITION ON 7 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	163	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%
Total	174	166	162	164	168	170	172		
Precision	95%	98%	98%	99%	92%	92%	98%		

from each other in *time domain*, our algorithm describes the human activity signal in a statistical way and gets rid of the temporal information in the data. Therefore, inverse processes could share lots of common features in *space domain*. As for sensor location classification, as illustrated in Table IV, most of the precision and recall are more than 98%. However, location e) *Left Thigh* and location f) *Right Ankle* get confused with each other the most. Specifically, the corresponding accuracy is around 92%. It indicates that the selected features described in Section III can not reliably distinguish the two cases. We could consider this issue to enhance the algorithm performance in the future work.

D. Comparison Between ℓ_1 and ℓ_2

As stated in Section II, ℓ_1 is a better heuristic for sparsity pursuit than ℓ_2 . As our last experiment, we want to empirically validate this point and compare the classification performance between ℓ_1 and ℓ_2 optimization strategies. As an example, Figure 4 shows the solutions from both ℓ_1 and ℓ_2 optimization with one test sample from activity 7 (kneeling right) at location d (right arm). As illustrated, the solution from ℓ_1 is quite sparse. 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 ℓ_2 spreads

out. The spikes are dense and distributed over all activity and sensor location classes.

Figure 5 illustrates the corresponding residual values between the test sample and all 98 classes defined by Eq.10 for both ℓ_1 and ℓ_2 . As shown, the class membership of the test sample can be easily told by the minimal residual (pointed by the red arrow) for ℓ_1 optimization strategy. For ℓ_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.

Finally, we compare the classification performance between ℓ_1 and ℓ_2 . Table V shows the results in terms of the recognition rates. As shown, ℓ_1 outperforms ℓ_2 across all three metrics consistently in terms of both recognition accuracy and stability. It is worth to emphasize that the enhancement from ℓ_1 compared to ℓ_2 has strong scalability: the larger the scale is, the more gain it has. Based on the observation in Figure 4, it is not surprised that ℓ_1 outperforms ℓ_2 overwhelmingly in terms of both accuracy and stability. Specifically, the co-recognition classification accuracy could be improved by 20.75% with ℓ_1 optimization. Correspondingly, the gain of stability from ℓ_1 optimization is 3.17X by average.

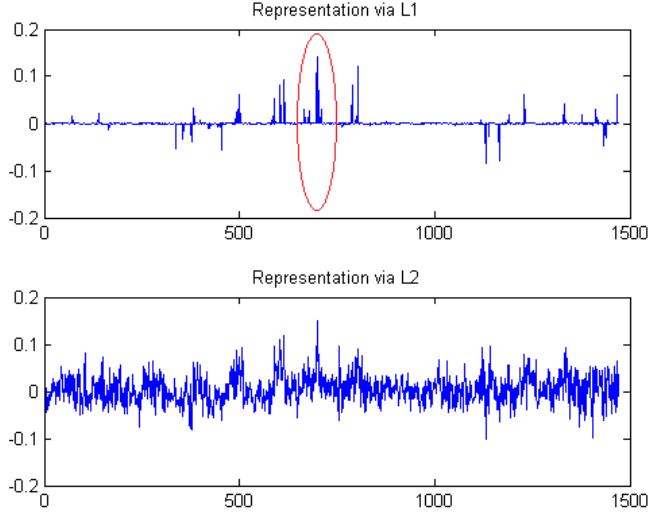


Figure 4. Solutions of ℓ_1 and ℓ_2 Optimization Strategies

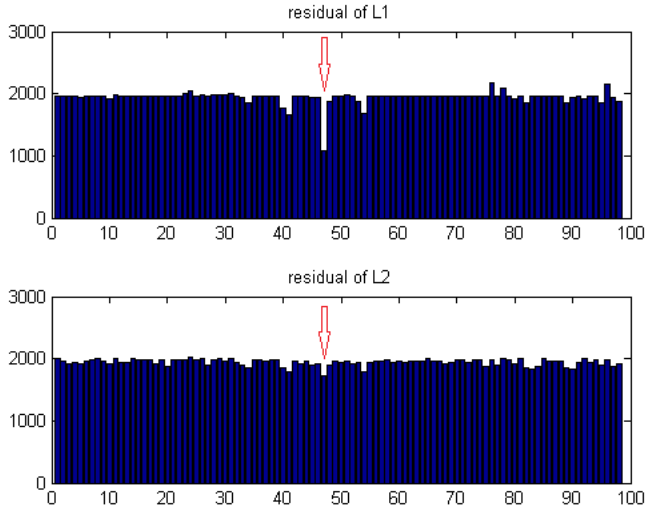


Figure 5. Residuals of 98 Classes of ℓ_1 and ℓ_2 Optimization Strategies

V. CONCLUSION

Inspired by the sparsity of human activity signal, we adopted the emerging compressed sensing theorem and proposed a novel framework to co-recognize human activity and sensor location in wearable sensor networks. The experimental results showed that our proposed method can achieve more than 87% recognition accuracy with 14 different activities and 7 on-body locations. Moreover, we also showed that the data presentation via ℓ_1 outperformed ℓ_2 in terms of both accuracy and robustness. Considering the promising results in the pilot study, we will consider to run the experiments with the large-scale group and evaluate more activities and sensor locations in the future work.

Table V
CLASSIFICATION PERFORMANCE COMPARISON OF ℓ_1 AND ℓ_2

	Activity&Location metric (1)		Activity metric (2)		Location metric (3)	
	mean	std	mean	std	mean	std
ℓ_1	87.72	1.26	89.00	1.13	96.13	0.41
ℓ_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

ACKNOWLEDGEMENT

The authors would like to thank Dr. Roozbeh Jafari, director of ESSP Lab in University of Texas at Dallas and Dr. Hassan Ghasemzadeh of Wireless Health Institute in UCLA for the experimental dataset and discussion.

REFERENCES

- [1] M. Zhang and A. A. Sawchuk, "Motion primitive-based human activity recognition using a bag-of-features approach," in *ACM SIGHIT International Health Informatics Symposium (IHI)*, (Miami, Florida, USA), pp. 1–10, January 2012.
- [2] N. Amini, M. Sarrafzadeh, A. Vahdatpour, and W. Xu, "Accelerometer-based on-body sensor localization for health and medical monitoring applications," *Pervasive and Mobile Computing*, vol. 7, pp. 746–760, 2011.
- [3] 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 Transactions on Biomedical Engineering*, vol. 50, pp. 711–723, 2003.
- [4] D. Donoho, "Compressed sensing," *IEEE Transactions on Information Theory*, vol. 52, pp. 1289–1306, 2006.
- [5] E. Candes, J. Romberg, and T. Tao, "Robust uncertainty principles: exact signal reconstruction from highly incomplete frequency information," *IEEE Transactions on Information Theory*, vol. 52, pp. 489–509, 2006.
- [6] B. Natarajan, "Sparse approximate solutions to linear systems," *SIAM Journal on Computing*, vol. 24, pp. 227–234, 1995.
- [7] A. Yang, R. Jafari, S. Sastry, and R. Bajcsy, "Distributed recognition of human actions using wearable motion sensor networks," *Journal of Ambient Intelligence and Smart Environments*, vol. 1, pp. 103–115, 2009.
- [8] C. Liu, Y. Yang, and Y. Chen, "Human action recognition using sparse representation," in *IEEE International conference on Intelligent Computing and Intelligent Systems*, (Shanghai, China), pp. 184–188, 2009.
- [9] M. Zhang and A. A. Sawchuk, "A feature selection-based framework for human activity recognition using wearable multimodal sensors," in *International Conference on Body Area Networks (BodyNets)*, (Beijing, China), November 2011.
- [10] Q. Shi, A. Eriksson, A. V. D. Hengel, and C. Shen, "Is face recognition really a compressive sensing problem?," in *IEEE International conference on Computer Vision and Pattern Recognition*, (Providence, RI, USA), pp. 1063–6919, 2011.