34th Annual International Conference of the IEEE EMBS
San Diego, California USA, 28 August - 1 September, 2012

Classification of Physical Activities Based on Sparse Representation

Shaopeng Liu, Robert X. Gao*, Dinesh John, John Staudenmayer, and Patty S. Freedson

Abstract—**This paper presents a new classification method for physical activity assessment, based on sparse representation. This method bypasses the need for feature extraction and selection that is typically involved for activity classification, and classifies activities using raw sensor signals directly. Higher discriminative power than that from the conventional k-nearest neighbor algorithm has been demonstrated through experiments performed on 105 subjects.**

I. Introduction

ACCURATE and reliable assessment of human physical activity (PA) is important for understanding individual behavior and quantifying the impact of PA on disease, as well as for examining determinants of PA in different populations. Physical activity by definition refers to bodily movement generated by skeletal muscle, and engaging in physical activities on a regular basis, such as walking, running, sports, household activities, is effective for improving human health, fitness and quality of life [1].

With the advancement of wearable electronics, body-worn activity monitors, consisting of either single or multiple accelerometers, or sensors of multiple modalities, have emerged as the technique of choice for PA assessment under free-living environments [2]. Combining machine learning or pattern recognition techniques to process these data can detect activity types and patterns of physical activity as well as intensity and energy expenditure [2].

Extracting and selecting the most relevant and informative features from the raw sensor signals has always been one of the typical and key problems in PA assessment and in machine learning/pattern recognition in general. An appropriate feature set is usually selected based on past experiences/literatures, or through certain feature selection algorithm performed on a large set of features. For example, in a previous study [2], a set of 63 features typically used in other PA studies, were pre-determined, and a subset of the features were then selected for minimum redundancy and maximum relevance by the mRMR algorithm [3]. Other feature selection algorithms, such as greedy forward search [4], floating search algorithm [5], etc., have also been used for feature selection. Such feature selection algorithms are suboptimal due to the requirement of a certain predefined "large" feature set. Since hundreds and thousands of features can be extracted out of a signal, it is almost impossible to select "optimal" feature set.

Recent development on the theory of sparse signal representation and compressive sensing [6], [7] has provided new stimulus to addressing the problem of feature selection and developing an alternative approach to classification, e.g., for face recognition [8] and fMRI data analysis [9]. In this presented study, a new classification method based on sparse representation for PA assessment is presented. Unlike the methods described in [10], the presented method *bypasses* the procedures for feature extraction and selection, and performs activity classification using raw sensor signals directly. The performance is experimentally evaluated by human subjects mimicking free-living activities in a laboratory setting.

II. Theoretical Framework

In this section, we exploit the theoretical foundation of sparse representation to perform activity classification. Different from using the features extracted from the raw sensor signals as in a typical activity classification procedure, a test signal to be classified will be represented in an overcomplete linear space where the base elements (basis) consist of a set of training sensor signals. Theoretically, human activities consist of complex data that constitute complex nonlinear models. Often times, such complex nonlinearity is approximated by linear models in a vector space with higher-dimensionality than that of a nonlinear model. One typical approach of performing the linear approximation is to map the original low-dimensional data into high dimensional space through some transformation or kernel functions [2], e.g., support vector machine (SVM) or artificial neural network (ANN). However, such approaches require the determination of an appropriate transformation or kernel function, as well the extraction of appropriate features.

Alternatively, that given sufficient training sensor signal samples (vectors) that span the activity vector space, it is possible to accurately represent a new test signal sample as a linear combination of the training samples [8]. Noteworthy is that it is a sparse representation that involves base elements (training samples) only from the same activity as the test sample. The concept of such approach is illustrated in Fig. 1. The left side of the equation is the test signal sample from an activity such as jogging, which is represented as a sparse linear combination (the right side of the equation) of training samples from different activities (middle of Fig. 1). The idea can be represented in matrix format, expressed as

$$y = Ax \tag{1}$$

Manuscript submitted on March 15, 2012. *Asterisk indicates corresponding author.*

Shaopeng Liu and Robert X. Gao* are with the Electromechanical Systems Laboratory, Department of Mechanical Engineering, University of Connecticut, Storrs, CT 06269, USA. (phone: 860-486-7110; fax: 860-486-5088; e-mail: rgao@engr.uconn.edu).

Dinesh John and Patty S. Freedson are with Physical Activity and Health Laboratory, Department of Kinesiology, University of Massachusetts Amherst, MA 01003, USA.

John Staudenmayer is with the Department of Mathematics and Statistics, University of Massachusetts Amherst, MA 01003, USA.

where y is the test sample, A is the training sample vector space, and x is the sparse coefficients.

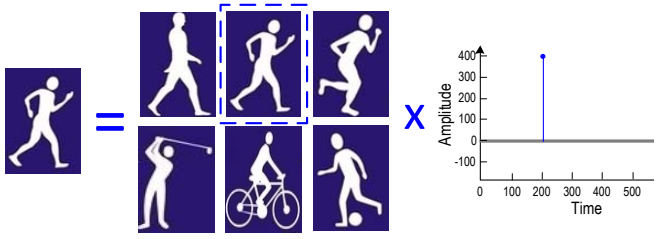

Fig. 1. Illustration of the sparse representation classification concept.

A. Sparse Representation Formulation

Typically, activity classification or any classification problem is to identify the activity type or class to which a new test sample (usually time-series signals measured by one or multiple sensors) belongs, by using labeled/known training samples from distinct activity types or classes. Assuming a total of K distinct classes (or activities), a training matrix that consists of n_i training samples from the ith class (or activity), can be constructed as $A_i = \left[v_{i,1}, v_{i,2}, \ldots, v_{i,n_i} \right] \in \Re^{m \times n_i}$, where the vector $v_{i,j} \in \Re^{m \times 1}, j = 1, 2, \ldots, n_i$ is the jth training sample with m raw sensor data points. For single sensor device, the training sample vector consists of data points from the sensor signal. While for multi-sensor inputs, each vector $v_{i,j}$ can be constructed by stacking the multi-sensor data together. Specifically, let $s^{(p)}_{i,j}$ denote the jth training samples of p sensors from the ith activity, and the multi-sensor training vector $v_{i,j}$ can thus be constructed as $v_{i,j} = [s^{(1)}{}^{T}_{i,j}, s^{(2)}{}^{T}_{i,j}, \ldots, s^{(p)}{}^{T}_{i,j}]^T$. The overall training sample matrix for the K classes, represented by A, is defined as

$$A \doteq \left[A_1, A_2, \ldots A_K \right] = \left[v_{1,1}, v_{1,2}, \ldots v_{1,n_1}, v_{2,1}, \ldots, v_{K,n_K} \right] \in \Re^{m \times n} \quad (2)$$

where n is the total number of training samples from the K classes, and $n = \sum_{i=1}^{K} n_i$.

For any test sample $y \in \Re^m$ that belongs to the ith class, given sufficient training samples from the ith class, y can thus be approximately represented as a linear span (or combination) of the training samples in A_i [8], expressed as

$$y = \alpha_{i,1} v_{i,1} + \alpha_{i,2} v_{i,2} + \ldots + \alpha_{i,n_i} v_{i,n_i} = A_i x_i \quad (3)$$

where $x_i = [\alpha_{i,1}, \alpha_{i,2}, \ldots, \alpha_{i,n_i}]^T$, is a set of n_i coefficients (scalars). The testing sample y can be rewritten in terms of the overall training sample matrix A as

$$y = A x_0 = \left[A_1, A_2, \cdots A_k \right] \cdot \left[x_1^T, x_2^T, \ldots, x_K^T \right]^T \quad (4)$$

where $x_0 = [0, \ldots, 0, \alpha_{i,1}, \alpha_{i,2}, \ldots, \alpha_{i,n_i}, 0, \ldots, 0]^T \in \Re^n$, is a global coefficient vector corresponding to the testing sample y. It is seen that except for the coefficients that are associated with the ith class, the rest are zero. Similarly, for a new test sample y' from the (i+1)th class, its corresponding global coefficient vector should have zero entries except for those associated with the (i+1)th class, specifically $\alpha_{i+1,1}, \alpha_{i+1,2}, \ldots, \alpha_{i+1,n_{i+1}}$.

It is further noted that such global coefficient vectors are discriminative in nature, and they uniquely represent test samples of different classes. Thus, the class (activity) to which a test sample y belongs to can be determined through its global coefficient vector, which can be obtained by solving the linear system of equations $y = Ax$ as in (4). Note that, for a test sample of a specific activity, the corresponding coefficient vector is sparse that only the coefficients associated with the activity are nonzero. As a result, the sparsest solution of (4) can be obtained through the following ℓ_0-minimization problem:

$$\tilde{x}_0 = \arg\min \|x\|_0 \quad \text{subject to } y = Ax \quad (5)$$

where $\tilde{x}_0$ is sparse solution, arg min is the argument of the minimum, $\| \cdot \|_0$ is the ℓ_0-norm, which yields the number of nonzero elements in a vector. Usually for physical activity classification, the system $y = Ax$ is underdetermined ($m << n$) due to the amount of the training samples needed, and finding the sparsest solution of such an underdetermined system of linear equations is thus NP-hard [8] such that it requires non-deterministic polynomial time.

With recent advancement in the field of compressive sensing and sparse representation [6], [7], it has been shown that the sparse solution of the ℓ_0-minimization problem is equivalent to the solution of the following ℓ_1-minimization problem when the solution x_0 is sparse enough [8]:

$$\tilde{x}_1 = \arg\min \|x\|_1 \quad \text{subject to } y = Ax \quad (6)$$

where $\tilde{x}_1$ is ℓ_1-minimization solution, $\| \bullet \|_1$ is the ℓ_1-norm, which yields the sum of the modulus of the elements in a vector. Solving ℓ_1-minimization is a developed problem, which can be solved by basis pursuit [11], and the problem can be rewritten as

$$\tilde{x}_1 = \arg\min \|x\|_1 \quad \text{subject to } \min\left(\|x\|_1 + \|y - Ax\|_2^2 \right) \quad (7)$$

B. Activity Classification via Sparse Representation

After obtaining the sparse solution via (7) for any given new test sample y, the next step is to identify the activity (class) to which y belongs to. An ideal sparse representation $\tilde{x}_1$ would consist of nonzero coefficients at the locations corresponding to the training matrix A_i of a specific activity, and the test sample y can then be easily assigned to that activity. However, the sparse solution in practice would be contaminated by noise or error due to the minimization process, such that $\tilde{x}_1$ would contain (usually relatively small) nonzero elements that correspond to multiple activities. Therefore, in order to establish a quantitative classification of the activity type, a test sample reconstruction procedure is developed: the test sample is reconstructed for each activity type as $\tilde{y}_i = A\delta_i(\tilde{x}_1)$, where $\delta_i(\tilde{x}_1)$ is a transformation function that replaces the elements in the sparse solution that correspond to the rest activity types rather than i with zeros. In

another word, the nonzero elements in $\delta_i(\tilde{x}_1)$ are only the elements in $\tilde{x}_1$ that are associated with the *i*th activity, such that $\delta_i(\tilde{x}_1) = [0,\ldots,0,x_i^T,0,\ldots,0]^T$. If the test sample *y* belongs to the *i*th activity, it can be expected that the residue, calculated as the root mean squared error between the given test sample *y* and the reconstructed $\tilde{y}_i$ is closest to zero (ideally zero) among all residues calculated for the *K* activities. The activity type of the test sample *y* can thus be classified as

$$k = \arg\min_i \left\| y - A\delta_i(\tilde{x}_1) \right\|_2 \tag{8}$$

Thus, such an activity classification algorithm based on sparse representation (SRC) using the original raw sensor signals instead of extracted features is summarized below:

Sparse Representation Classification Algorithm

1. Construct a global training sample matrix: $A = (A_1, A_2, ..., A_k) = (v_{1,1}, v_{1,2}, ..., v_{k,n_k}) \in \Re^{m \times n}$.
2. Normalize the columns of ***A*** to avoid samples in greater numeric ranges dominate those in smaller ranges.
3. Given a test sample signal ***y***, calculate the sparse solution $\tilde{x}_1 = \arg\min \|x\|_1$, subject to $y = Ax$ or $\min\left(\|x\|_1 + \|y - Ax\|_2^2\right)$.
4. Project the sparse coefficients $\tilde{x}_1$ onto each sub-training space $\boldsymbol{A_i}$, $\delta_i(\tilde{x}_1) = [0,\ldots,0,x_i^T,0,\ldots,0]^T$, x_i is the coefficients in $\tilde{x}_1$ corresponding to *i*th activity.
5. Calculate the residue for each activity $\left\| y - A\delta_i(\tilde{x}_1) \right\|_2$.
6. Classify ***y***: $k = \arg\min_i \left\| y - A\delta_i(\tilde{x}_1) \right\|_2$.

C. Random Feature Projection

When the classification system have a large number of sensors and/or when individual training sample consists of a substantial amount of sensor data points, the training sample matrix would therefore be constructed with a high dimensionality of *m*. For example, for 10-channel sensor signals sampled at 1 kHz, if individual training sample length is 10 second, the dimension of the constructed training matrix is in the order of 10^5. Such a high dimensionality would increase the computational cost substantially.

Construct a transformation matrix $R \in \Re^{d \times m}$ with $d << m$, and each element of *R* is independently sampled from a standard Gaussian distribution with the mean equal to 0 and variance equal to 1. Such a matrix *R* is considered to be a random feature projection that extracts random "features" from the original raw sensor signals. Multiplying *R* to both sides of (1), the dimensionality of A is then reduced from *m* to *d*, and (1) can thus be rewritten as

$$Ry = RAx_0 \tag{9}$$

Denote $\hat{y} \doteq Ry$, $\hat{y} \in \Re^{d \times 1}$, and the sparse solution $\tilde{x}_1$ can still be obtained by solving the following reduced ℓ_1-minimization problem [8]:

$$\begin{aligned} \tilde{x}_1 = \arg\min \|x\|_1, \quad & \text{subject to} \quad \hat{y} = RAx \\ & \text{or subject to} \quad \min\left(\|x\|_1 + \|\hat{y} - RAx\|_2^2\right) \end{aligned} \tag{10}$$

Furthermore, it is reported in previous study [8] that if the solution x_0 is sparse enough such that the number of the nonzero elements in x_0 is far less than the number of the total training samples in *A*, the sparse solution $\tilde{x}_1$ can be correctly obtained from any sufficiently large number of *d* of the projection *R*. Thus, the SRC algorithm can be updated by replacing the training sample matrix *A* with *RA*, and test sample *y* with *Ry*.

III. Experiments

The performance of the sparse representation based activity classification algorithm has been evaluated through experiments, where a total of 105 healthy subjects (44 male, 61 female) were recruited. Specifically, the group has the following characteristics (mean ± standard deviation): age = 34.7±14.2 years, weight = 68.7±17.1 kg, height = 166.8±11.3 cm, and body mass index = 24.4±4.5 kg/m^2. Each subject wears a multi-sensor system [2] that consists of two tri-axial accelerometers (MMA7260QT, Freescale) placed at the hip and wrist, and one ventilation sensor (1325 Piezo Crystal Sensor, Ambu Sleepmate) secured to the abdomen (AB) at the level of umbilicus of the test subjects. The accelerometers measure the trunk and arm motions, and the ventilation sensor measures the expansion and contraction associated with breathing rate and volume representing the physiological response to bodily movement. Data from the three sensors are sampled at 30 Hz and pre-processed by a microcontroller on board a data logger, worn at the subject's waist and subsequently stored into a micro-flash memory.

Each subject completed activities in one of the two following groups in a random order: 1) computer work (CW), moving boxes (MB), cycling with 1-kp resistance (C1), treadmill walking at 3.0 mph with 0% grade (T3-0), treadmill walking at 3.0 mph with 5% grade (T3-5), treadmill walking at 4.0 mph with 5% grade (T4-5), and tennis (TE); and 2) filing papers (FP), vacuuming (VA), self-paced walk (SW), cycling with 2-kp resistance (C2), treadmill walking at 4.0 mph with 0% grade (T4-0), treadmill jogging at 6.0 mph and 0% grade (T6-0), and basketball (BA). Subjects were not required to complete all activities and were allowed to skip the high-intensity activities that they did not feel comfortable to complete, e.g., treadmill at 6.0 mph, basketball or tennis. Each activity, when performed, lasted for 7 minutes long, followed by a 2-minute rest period.

Subject data of each sensor were also divided into 5-second intervals to construct the training matrix for the SRC algorithm. Same data set was also processed by a conventional technique – k-nearest neighbor (kNN) for comparison. Leave-one-out cross validation was performed during the process.

IV. Results

Figure 2 shows an example of the sparse representation based classification of a test sample from the activity CW, among 6 different activities. Each activity had 80 training samples, which constructed a global training sample of 480

training vectors from the 6 activity types. It is seen from Fig. 2(a) that the obtained sparse solution has more nonzero elements corresponding to CW than the rest activities, and from the quantitative calculation of the residues as shown in Fig. 2(b) that the residue corresponding to the CW activity was the smallest and closest to zero among the 6 activities. Therefore, the activity type of the given test sample was identified as activity CW.

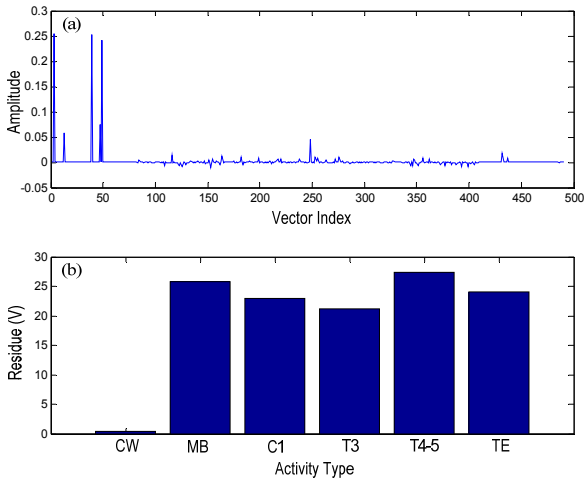

Fig. 2. An example of sparse representation classification. (a) Sparse solution of a test sample from the activity CW; (b) Residues of the reconstructed test sample of each activity type.

As seen in section II.A, while using the random feature projection may reduce the computational complexity, it may also affect the activity classification performance without choosing an appropriate dimension of the projection matrix *R*. To evaluate the effect of the dimension of *R*, a preliminary test has been performed as shown in Fig. 3, where the activity classification accuracy was calculated for each matrix dimensionality ranging from 10 to 150. It is seen that the classification rate increases as the dimension of the projection increases. When the dimension of the project exceeds 50, the increase of the classification rate reaches a plateau, and the activity classification accuracy reaches its maximum of 83.0% at the dimension of 60. Thus, a 60-dimension random feature projection would be reasonable for activity classification.

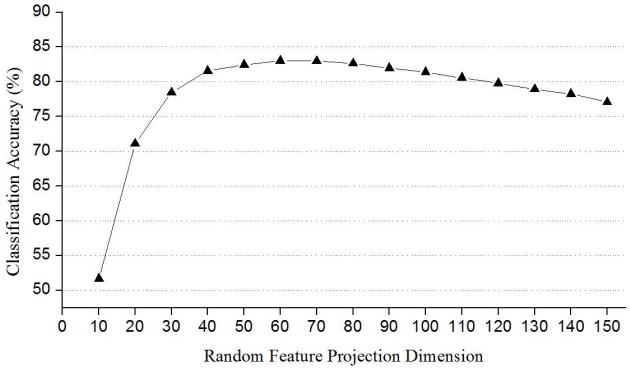

Fig. 3. Classification rates corresponding to different projection dimensions.

With the random feature projection, as long as the sparse solution can be correctly obtained [8], the developed sparse representation classification algorithm will always yield the same classification results as using the original training sample matrix of large dimensionality. Furthermore, by applying such a random feature projection, it can be concluded that the selection of an appropriate feature set is no longer critical to the classification problem, since a random feature extraction can also achieve the same performance.

Table I compares the activity classification accuracy, mean and standard deviation (SD) of the 105 subjects by the SRC algorithm with that by using the k-Nearest Neighbor (kNN) algorithm (with mean, standard deviation, and 10th, 25th, 50th, 75th, and 90th percentiles extracted as features [2]). The result has demonstrated better discriminative performance of the SRC than kNN.

TABLE I
COMPARISON OF ACTIVITY CLASSIFICATION ACCURACY

Methods	SRC	kNN
Classification Accuracy (Mean ± SD)	83.0 ± 5.6	72.7 ± 16.8

V. CONCLUSION

A new classification method based on sparse representation of data has been developed for physical activity assessment. Experimental results have demonstrated higher classification accuracy by the SRC algorithm than that by conventional classification methods such as kNN. Future study will comparatively evaluate the performance of the developed algorithm against other activity classifiers, as well as perform a complexity analysis of the developed algorithm in terms of its implementability for portable activity monitoring devices.

REFERENCES

[1] W. D. McArdle, F. L. Katch, and W. L. Katch, *Exercise Physiology: Nutrition, Energy, and Human Performance* (5th Edition), Lippincott Williams & Wilkins, p. 871, 2001.

[2] S. Liu, R. Gao, D. John, J. Staudenmayer, and P. Freedson, "Multi-sensor data fusion for physical activity assessment", *IEEE Transactions on Biomedical Engineering*, vol. 59, No. 3, pp. 687-696, March 2012.

[3] H. Peng, F. Long, and C. Ding, "Feature selection based on mutual information: criteria of max-dependency, max-relevance, and min-redundancy," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 27, pp.1226-1238, 2005.

[4] P. A. Devijver and J. Kittler, *Pattern recognition: A statistical approach*. Englewood Cliffs: Prentice Hall, 1982.

[5] P. Pudil, J. Novovičová, J. Kittler, "Floating search methods in feature selection," *Pattern Recognition Letters*, vol. 15, pp. 1119-1125, November, 1994.

[6] D. Donoho, "For most large underdetermined systems of linear equations the minimal *l*1-norm solution is also the sparsest solution," *Communications on Pure and Applied Mathematics*, vol. 59, no. 6, pp. 797-829, 2006.

[7] E. Candès and T. Tao, "Near-optimal signal recovery from random projections: universal encoding strategies?" *IEEE Transactions on Information Theory*, vol. 52, no. 12, pp. 5406-5425, 2006.

[8] J. Wright, A. Y. Yang, A. Ganesh, S. S. Sastry, and Y. Ma, "Robust face recognition via sparse representation," *IEEE Transactions on Pattern Analysis and Machine*, vol. 31, 2009, pp. 210-227.

[9] Y. Li, P. Namburi, Z. Yu, C. Guan, J. Feng, and Z. Gu, "Voxel Selection in fMRI data analysis based on sparse representation", *IEEE Transactions on Biomedical Engineering*, vol. 56, pp. 2439-2451, Oct. 2009.

[10] T. Mitchell, *Machine Learning*. McGraw-Hill, 1997.

[11] S. Chen, D. Donoho, and M. Saunders, "Atomic decomposition by basis pursuit," *SIAM Review*, vol. 43, no. 1, pp. 129-159, 2001.