Activity Recognition via Distributed Random Projection and Joint Sparse Representation in Body Sensor Networks

Ling Xiao*, Renfa Li, Juan Luo, and Mengqin Duan

Laboratory of Embedded Systems and Networks,
Hunan University, Changsha 410082, China
{xiaoling,juanluo}@hnu.edu.cn, lirenfa@vip.sina.cn

Abstract. Designing power-aware signal processing algorithms for activity recognition is challenging as special care needs to be taken to maintain acceptable classification accuracy while minimizing the energy consumption. This paper utilizes the theory of distributed random projection and joint sparse representation to develop a simultaneous dimension reduction and classification approach for multi-sensor activity recognition in BSNs. Both temporal and spatial correlations of sensing data among the multiple sensors are exploited for the purpose of compression and classification. Activity recognition with multiple sensors is formulated as a multi-task joint sparse representation model to combine the strength of multiple sensors for improving the classification accuracy. This method is validated on the WARD dataset using inertial sensors placed on various locations on a human body. Experimental result shows that the proposed DRP-JSR approach achieves better classification performance that is competitive with traditional classifier.

Keywords: Activity Recognition, Joint Sparse Representation, Random Projection, Body Sensor Network.

1 Introduction

Wireless body sensor networks (BSN) with multiple inertial sensors are widely used in various studies on human body movement[1]. A lot of pattern recognition and machine learning algorithms were developed to model and recognize human activities. As for the recognition techniques, a large number of classification methods have been investigated [2]. Some studies incorporated the idea of simple heuristic classifier, whereas others employed more generic and automatic methods from the machine learning literature including the decision trees , nearest neighbor (NN), Bayesian networks, support vector machines (SVM), Artificial neural networks (ANN) and Hidden Markov Model (HMM). A particular interest of multi-sensor fusion is classification, where the ultimate question is how

* This work was partially supported by Program for New Century Excellent Talents in University(NCET-12-0164), Natural Science Foundation of Hunan(13JJ1014).

L. Sun, H. Ma, and F. Hong (Eds.): CWSN 2013, CCIS 418, pp. 51–60, 2014.
© Springer-Verlag Berlin Heidelberg 2014

to take advantage of having related information from different sensors recording the same physical event to achieve improved classification performance. In activity recognition of multi-sensor, a final decision can be made using either a data fusion or a decision fusion scheme. In the data fusion, features from all sensor nodes are fed into a central classifier. The classifier then combines the features to form a higher dimensional feature space and classifiers movements using the obtained features [3]. In the decision fusion, however, each sensor node makes a local classification and transmits the result to a central classifier where a final decision is made according to the received labels [4]. However, most existing techniques are designed for single observation based classification, which are clearly not optimal due to the failure of exploiting the correlations among the multiple observations of the same physical object.

However, the battery limitations of the BSN severely limit the maximum deployment time for continuously monitoring human body. This problem is often solved by shifting some processing to the local sensor nodes to reduce a very heavy communication cost. Designing power-aware signal processing algorithms for activity recognition is challenging as special care needs to be taken to maintain acceptable classification accuracy while minimizing the energy consumption. This paper focuses on developing a computationally simple and energy-efficiency algorithm for action recognition. Compressive sensing (CS) [5] is new method for recovering of sparse or compressible signals from a small set of non-adaptive, linear measurements. It has been shown that random projections (RP) are a near-optimal measurement scheme. Distributed compressive sensing (DCS) [6] is an extension of the CS acquisition framework to correlated signal ensembles. Using DCS as the data acquisition approach in BSNs can significantly reduce the energy consumed in the process of sampling and transmission through the network, and also lower the wireless bandwidth requirements for communication.

Sparsity has been the key factor of compressive sensing and has been playing an important role in many fields. Sparse signal representations from over-complete dictionaries have far-reaching significance in signal processing. Recently, a sparse representation classification (SRC) method for face images is developed in [7], this work has shown that the sparse coefficients are also discriminative. The term Joint Sparsity was first coined in [6]. Yuan et al. [8] investigated the problem of multi-task joint sparse representation and classification and its applications to visual recognition. Zhang et al. [9] proposed a joint sparse representation for multi-view face recognition.

The goals and contributions of this paper are as follows: 1) to combine distributed random projection with joint sparse representation for power-aware classification; 2) to extend the sparsity-based classification approach to handle the multi-sensor classification problem with joint-structured-sparsity priors; 3) to propose the DRP-JSR classification algorithm and to find a sparse Bayesian learning algorithm for solving this problem.

The rest of the paper is organized as follows: In Section 2, we define the problem of activity recognition and review the sparse representation classification method. In Section 3, we present the proposed distributed random projection and joint sparse representation classification for multi-sensor. In Section 4, we evaluate the efficacy of the proposed method under various compression ratios. Finally, we make some discussions and conclude this paper in Section 5.

2 Problem Descriptions

2.1 Problem Definition

Beginning with the problem formulation, let there be a set of J wearable sensors, each with consisting of a 3-axis accelerometer *(x,y,z)* and a 2-axis gyroscope (θ, φ), attached to the human body. Then, let

$$a^j(t) = (x^j(t), y^j(t), z^j(t), \theta^j(t), \varphi^j(t),) \in \mathbb{R}^5 \tag{1}$$

denotes the 5 measurements provided by node j at time t, and

$$v^j = [a^j(1), a^j(2), \cdots, a^j(h)]^{\mathrm{T}} \in \mathbb{R}^{5h} \tag{2}$$

corresponds to an action segment of length h by node j.

Consider a multi-task (multi-sensor) C-class classification problem. Suppose we have a training set of n samples in which each sample was collected by J different sensors. For each sensor $j = 1, \cdots, J$, we denote $V^j = [V_1^j, V_2^j, \cdots, V_C^j]$ as a $N \times n$ training feature matrix in which $V_i^j = [V_{i,1}^j, V_{i,2}^j, \cdots, V_{i,n_i}^j] \in \mathbb{R}^{N \times n_i}$ with respect to C classes. Here, each sub-dictionary represents a set of training data from the jth sensor labeled with ith class. Accordingly, , which we usually call an atom in the dictionary is the kth training sample for jth sensor and ith class. In addition, we have a training label vector $L \in \mathbb{R}^n$ associated with v. Notice that n_i is the number of training sample for class ith and N is the feature dimension of each sample ($N = 5h$), therefore, the total samples is $n = \sum_{i=1}^{C} n_i$. Given a test sample v_{test} collected by J sensors $\{v_{test}^1, v_{test}^2, \cdots, v_{test}^J\}$, we want to decide which class the sample v_{test} belongs to. This can be formally represented as

$$\hat{i} = argminE(V, L, v_{test}) \tag{3}$$

Where $E(V, L, v_{test})$ is a cost function defining the classification problem.

2.2 Sparse Representation-Based Classification

We first review the single task (single sensor) sparse representation based classification method. A SRC method for single image based face recognition has been proposed in [7]. This method casts the task of face recognition as one of classifying between linear regression models via sparse representation. The sparsest linear combination of a test face image is sought using all the training images, and the dominant sparse coefficients reveal its identity.

A single new test sample v_{test}^{j} (i.e., J=1) can be represented in terms of the atoms form a structured dictionary V^{j} as follows:

$$v_{test}^{j} = V^{j}\alpha^{j} + \varepsilon \tag{4}$$

where $\alpha^{j} = [0, \cdots 0, \alpha_{k,1}, \alpha_{k,2}, \cdots, \alpha_{k,n_i}, 0, \cdots 0]^{\mathrm{T}} \in \mathbb{R}^{n}$, α^{j} is a sparse representation vector whose entries are zero except those associated with the same class as v_{test}^{j}.

The theory of sparse representation and compressive sensing reveals that the solutions exist to uniquely recover sparse solution α^{j} via l_1-minimization:

$$\hat{i} = argmin \left\|\alpha\right\|_1 \, subject \; to \left\|v_{test}^{j} - V^{j}\alpha^{j}\right\|_2 \leq \varepsilon \tag{5}$$

After recovering the sparse representation vector α^{j} , the class label for $v_{k,test}$ is assigned to the class with the smallest residual

$$\hat{i} = SRC(v_{test}^{j}) = argmin \left\|v_{test}^{j} - V^{j}\delta_{i}^{j}(\hat{\alpha})\right\|_2 \tag{6}$$

Where $\delta_{i}^{j}(\hat{\alpha})$ is defined as a vector indicator function, keeping the coefficients corresponding to the ith class while setting all others to be zero.

In this paper, we extend SRC to handle the multi-sensor classification problem with joint-structured-sparsity priors. It is important to note that Human body motions usually exhibit a high degree of coherence and correlation in patterns. By temporal correlation, body signs sensed by a single node typically change smoothly and slowly. By spatial correlation, body signs measured on different nodes are typically correlated because body components are connected and they normally move with certain rhymes. On the other hand, sensors of the same node may also exhibit strong spatial correlations, especially among the three axes of a triaxial accelerometer. We call such spatial correlations intra-node spatial correlations and those among sensors of different nodes inter-node spatial correlations. By exploiting the structural information across the multiple sparse representation vectors for the multiple sensors, we can reduce the number of measurements required for proper model estimation and improve the accuracy of classification.

Three different models are investigated in this paper. 1) Separate sparse representation and their sparse patterns may be quite different. 2) All the representation vectors have equal values. 3) Sparse coefficient vectors share similar patterns, but with different coefficient values.

3 DRP-JSR Classification

3.1 Distributed Random Projection

Random projections of the signal measurements are performed at each source node, only taking into account the temporal correlation of the sensor readings.

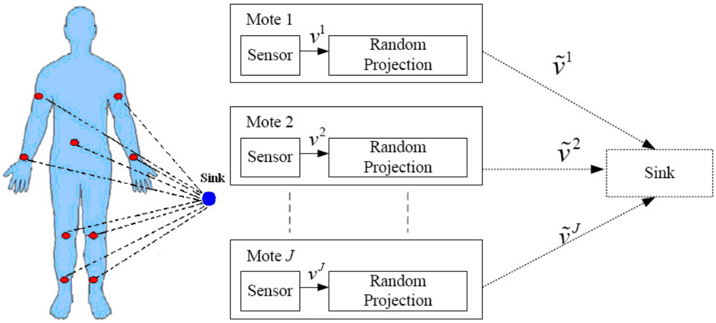

Fig. 1. Framework of distributed random projection

We denote by Φ_j the measurement matrix for sensor j; Φ_j is $M \times N$ and, in general, the entries of Φ_j are different for each j. After random projection matrix Φ_j is chosen on each node j, there are:

$$\tilde{v}^j = \Phi_j v^j \tag{7}$$

Where $\tilde{v}^j$ is a vector after RP.

Gaussian and Bernoulli random matrix have been proven to follow RIP with a very high probability [10].

After RP, typically the feature dimension M is much smaller than N. Furthermore, by exploiting the relationships between these J sets of sensor measurements, the number of transmissions to the sink can be further reduced, with a consequent reduction of the energy consumed by the sensor nodes.

3.2 Classification by Joint Sparse Representation

Every sensor sends random projection vector $\tilde{v}^j$ to the base station.

The spatial correlation is then exploited at the sink by means of suitable decoders through a joint sparsity model able to characterize different types of signals. For each sensor, there is equation:

$$\tilde{v}^j_{test} = \Phi_j V^j \beta^j + \varepsilon_j \tag{8}$$

To handle multi-sensor classification, the simplest idea would be to perform SRC method for each sensor separately and the recovered spare representation vectors may be quite different, as shown by a graphical illustration in figure.3 (a). The final decision is based on the lowest total reconstruction error accumulated from all the sensors, which is called by SRC-S.

It is clear that the SRC-S approaches do not exploit the relationship between different sensors except at the post-processing step where decision is made via fusion. Rather than doing post-processing during the decision fusion, it is more

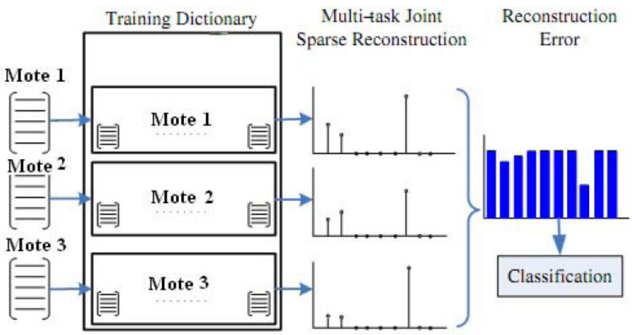

Fig. 2. Scheme illustration of activity recognition with multi-sensor joint sparse representation

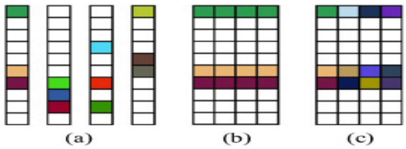

Fig. 3. Pictorial illustration of different sparsity models for coefficient matrix

robust to recover the sparse representation vectors for the J sensors simultaneously. This can be done by exploiting the correlations among measurements of multiple sensors during the sparse representation process, and then make a single decision based on the J sparse representation vectors jointly. In the following, we exploit the correlations among the multiple measurements of multi-sensor by imposing different joint-structured-sparsity constraints on their sparse coefficient vector.

The most direct method to enforce joint structures on the multiple sparse representation vectors would be making the assumption that all the measurements would have the same sparse representation vector with respect to a dictionary, as shown by a graphical illustration in figure.3 (a), which is called SRC-A. The base station collects J sensors random projection vector and constitutes as $\tilde{v}$:

$$\tilde{v} = [\tilde{v}^1, \cdots, \tilde{v}^J]^{\mathrm{T}} = \varPhi v \tag{9}$$

Where $\varPhi \in \mathbb{R}^{MJ \times NJ}$ is a block-diagonal matrix which constitutes by J nodes random projection matrices.

Correspondingly, we form a large concatenated dictionary as

$$V = [{V^1}^{\mathrm{T}}, \cdots, {V^J}^{\mathrm{T}}]^{\mathrm{T}} \tag{10}$$

where each V^j can be constructed by the training samples from the corresponding jth sensor. The sparse representation can be formulated by equation (9), we enforce all the sparse representation vector is the same, i.e., $B = \beta^1 = \beta^1 = \cdots = \beta^J \in \mathbb{R}^N$.

$$\tilde{v}_{test} = \Phi V \beta + \varepsilon \tag{11}$$

Although the SRC-A can reduce the degrees of freedom greatly, the constraint that all the measurements would have the same sparse representation vector is a restrictive, which is often violated in typical sparsity-based classification problems.

Therefore, we make a relax assumption by assuming that the sparse representation vector for multi-sensor share the location of nonzero coefficients, but the values of the coefficients may be different for different sensors, as shown in figure 3.(c). The joint sparse representation can be represented by equation (12), which is named DRP-JSR.

$$\tilde{v}_{test}^1 = \Phi_1 V^1 \beta^1 + \varepsilon_1, \cdots, \tilde{v}_{test}^J = \Phi_J V^J \beta^J + \varepsilon_J \tag{12}$$

The rationale behind the DRP-JSR method is that the multiple measurements are highly correlated; thus, they tend to be represented by the same set of atoms. However, as aforementioned, due to the variation of environment and imperfection of the measurement process, all the J sensors are not exactly the same. Therefore, it is more reasonable to represent them with respect to the same set of atoms but weigh them with different coefficient values.

In order to seek for this row-sparse matrix with common sparse support as equation (12), an efficient algorithm for simultaneous sparse linear-regression of multiple related signals is required to that take into account this precious piece of sparsity as a prior information. We found the algorithms based on hierarchical Bayesian model [11] for this problem. A shared prior is placed across all of the J sensors. Under this hierarchical Bayesian modeling, data from all J sensors contribute toward inferring a posterior on the hyerparameters, and once the shared prior is thereby inferred, the data from each of the J individual sensors is then employed to estimate the sensor-dependent sparse coefficients. In the practice of activity recognition, when the sensors to be learned share some common latent factors, it may be beneficial to take into this relation into account. When applied to sparse learning, joint sparsity is always taken into account for multi-task learning.

After recovering the spare representation coefficient matrix, we can estimate the class label. Similar to the single task case, where minimal reconstruction residual vector criteria are used, we make a decision based on the lowest total reconstruction error accumulated from all the sensors.

$$\hat{i} = argmin \sum_{j=1}^{J} \left\| \tilde{v}_{test}^j - \tilde{V}^j \delta_i^j (\hat{\beta})^j \right\|_2 \tag{13}$$

Where $\delta_i(.)$ denotes the operation of preserving the rows of matrix $\hat{\beta}^j$ corresponding to class i and setting all others to be zero.

4 Experiment Results

4.1 WARD Dataset

For this work, we used the wearable action recognition database (WARD), which is implemented by Yang et al. of University of California, Berkeley[12]. The five sensor nodes (J=5) contain a 3-axis accelerometer and a 2-axis gyroscope, placed at different locations on the bodyone on the waist, two on the wrists, and two on the ankles. The sampling rates for both accelerometer and gyroscope are set to 20Hz. The dataset has been made publicly available, and consists of data recorded from 20 participants with different gender, age, height, and weight, for 13action categories: 1. Stand (ST). 2. Sit (SI). 3. Lie down (LI). 4. Walk forward (WF). 5. Walk left-circle (WL). 6. Walk right-circle (WR). 7. Turn left (TL). 8. Turn right (TR). 9. Go upstairs (UP). 10. Go downstairs (DO). 11. Jog (JO). 12. Jump (JU). 13. Push wheelchair (PU). Each participant performs five trials for each action. In total, there are 1300 examples.

We currently process all data offline in MATLAB. Our experiments use Sparse Bayesian Learning toolbox[10]. For each motion sequence in the WARD database, we randomly sample an action segment of length h in time. In the experiment, we choose h=40 as a short action duration, which corresponds to 2 seconds given the 20Hz sampling rate. The raw sensor sampled data are filtered using a five-point moving average to reduce high frequency noise. The training sets and the test sets are designed as follows. During the test, we employ a leave-one-subject-out validation approach to examine the subject-independence classification performance on a test sequence. This validation process was repeated for all twenty subjects. The number of samples in training sets from 19 subjects is 1235, i.e. n=1235.

4.2 Recognition Accuracy under RP

The Gaussian random projection matrices were chosen in our experiments. For each kind of random matrices, the validation process was repeated 10 times, a group of RP matrices were generated randomly every time. Each group of RP matrices consists of five random matrices Φ_j, in which each matrix corresponds to one sensor node respectively.

To investigate the robustness of the proposed SRC-RP algorithm for dimensionality reduction by RP, we set five different compression ratios of 0, 0.3, 0.5, 0.7 and 0.9. Compression ratio is defined as $(N\text{-}M)/N$. Table 1 gives the mean and standard deviation for recognition accuracies of the three approach with various compression ratios, the first column indicates the ratio of compression.

As shown in Table 1, with the increase of compression ratios, the mean recognition accuracies of three classifiers degrade slightly while the standard deviations of the recognition accuracies arise. It is observed that DRP-JSR method outperform the other method under the same compression ratio, which shows the superior classification performance of the joint-sparsity-based classification methods.

Table 1. Activity recognition accuracies (meanstandard ± deviation)of the three different sparsity models

Compression ratio	DRP-JSR	SRC-S	SRC-A
0	88.77	85.46	83
0.3	87.50±2.01	84.60±2.36	83.55±2.47
0.5	87.22±2.16	84.27±2.64	83.08±2.48
0.7	87.05±2.19	83.94±2.73	82.94±2.49
0.9	83.74±3.25	79.56±3.43	80.45±3.76

4.3 Compared with other Classification Methods

To further validate the performance of the proposed method, four common classifiers were carried out to recognize human activities on the WARD. The classifiers were nearest neighbor (NN) classifier, nearest subspace (NS), Bayesian Networks, and SVM. Bayesian Network was implemented under WEKA environment, and SVM was implemented under MATLAB environment by using LIBSVM toolbox. The average recognition accuracies of subject-independent for three classifiers are given in Table 2. As listed in Table 2, the DRP-JSR method achieved higher accuracy under the same compression ratio.

Table 2. Activity recognition accuracies (meanstandard ± deviation)of the common classifiers

Compression ratio	DRP-JSR	NS	NN	Bayes Net	SVM
0	88.77	83.2	81.0	87.8	87.9
0.3	87.50±2.01	82.85±1.66	79.38±1.89	83.22±1.66	83.29±1.65
0.5	87.22±2.16	82.20±1.83	79.27±2.18	83.57±1.59	83.02±1.52
0.7	87.05±2.19	81.66±2.13	78.51±2.58	82.05±1.36	82.51±1.58
0.9	83.74±3.25	78.75±5.29	76.40±3.61	77.87±1.83	80.01±1.63

5 Conclusion

In this paper, we generalize SRC to handle the multi-sensor classification problem with distributed random projection and joint sparsity priors. By exploiting the structural information across the multiple sparse representation vectors for the multiple sensors, we can reduce the number of measurements required for proper model estimation and improve the accuracy of classification. Several different models are investigated in this paper. 1) Separate sparse representation and their sparse patterns may be quite different. 2) All the representation vectors have equal values. 3) sparse coefficient vectors share similar patterns (selecting the same set of atoms), but with different coefficient values. We exploit the data correlation both temporally and spatially. The projections of the signal measurements are performed at each source node, only taking into account the temporal correlation of the sensor readings. The spatial correlation is then exploited at the sink through a joint sparsity model able to characterize different types of signals.

In the future, we would like to progress this work. With the development of compressive sensing and solution, we will find the analytical relationship between the RIP constants and the action recognition results and give a strong foundation to decide the number of projections required to get robust recognition results.

References

1. Seeger, C., Buchmann, A., Laerhoven, K.V.: MyHealthAssistant: A phone-based Body sensor network that captures the wearer's exercises throughout the Day. In: 6th IEEE International Conference on Body Area Networks, pp. 181–189. IEEE Press, New York (2011)
2. Yang, C.C., Hsu, Y.L.: A review of acclerometry-based wearable motion detector for physical activity monitoring. Sensors 10, 7772–7788 (2010)
3. Reichardt, D., Miglietta, M., Moretti, L., Morsink, P., Schulz, W.C.: CarTALK 2000: safe and comfortable driving based upon inter-vehicle-communication. In: IEEE Intelligent Vehicle Symposium, pp. 545–550. IEEE Press, Piscataway (2002)
4. Ghasemzadeh, H., Jarari, R.: Physical movement monitoring using body sensor networks: a phonological approach to construct spatial decision trees. IEEE Transactions on Industrial Informatics 7(1), 66–77 (2011)
5. Donoho, D.L.: Compressed sensing. IEEE Trans. on Information Theory 52(4), 1289–1306 (2006)
6. Barson, D., Wakin, M.B., Duate, M.F., Sarvotham, S., Baraniuk, R.G.: Distributed compressed sensing (November 2005) (preprint)
7. Wright, J., Yang, A., Ganesh, A., Sastry, S.S., Ma, Y.: Robust face recognition via sparse representation. IEEE Trans. on Pattern Anal. Mach. Intell. 31(2), 210–227 (2009)
8. Yuan, X.P., Liu, X., Yan, S.: Visual classification with multitask joint sparse representation. IEEE Trans. on Image Processing 21(10), 4349–4360 (2012)
9. Zhang, H., Nasrabadi, N.M., Zhang, Y., Huang, T.S.: Joint dynamic sparse representation for multi-view face recognition. Pattern Recognition 45, 1290–1298 (2012)
10. Baraniuk, R., Davenport, M., Devore, R., Wakin, M.: A simple proof of the restricted isometry property for random matrices. Constr. Approx. 28(3), 253–263 (2008)
11. Ji, S., Dunson, D., Carin, L.: Multitask compressive sensing. IEEE Trans. on Signal Processing 57(1), 92–106 (2009)
12. Yang, A., Kuryloski, P., Bajcsy, R.: WARD: a wearable action recognition database (2009), `http://www.eecs.berkeley.edu/~yang/software/WAR/`