

# COMPRESSED CLASSIFICATION OF OBSERVATION SETS WITH LINEAR SUBSPACE EMBEDDINGS

*Dorina Thanou and Pascal Frossard*

Signal Processing Laboratory (LTS4)  
Ecole Polytechnique Fédérale de Lausanne (EPFL)  
CH- 1015 Lausanne, Switzerland  
Email: {dorina.thanou, pascal.frossard}@epfl.ch

## ABSTRACT

We consider the problem of classification of a pattern from multiple compressed observations that are collected in a sensor network. In particular, we exploit the properties of random projections in generic sensor devices and we take some first steps in introducing linear dimensionality reduction techniques in the compressed domain. We design a classification framework that consists in embedding the low dimensional classification space given by classical linear dimensionality reduction techniques in the compressed domain. The measurements of the multiple observations are then projected onto the new classification subspace and are finally aggregated in order to reach a classification decision. Simulation results verify the effectiveness of our scheme and illustrate that compressed measurements combined with information diversity lead to efficient dimensionality reduction in simple sensing architectures.

**Index Terms**— Random projections, dimensionality reduction, multiple observations.

## 1. INTRODUCTION

The tremendous increase in the amount of data captured by different sensing systems has stimulated a lot of research toward the design of effective methods that could identify the relevant information in complex data. Due to this redundancy, the problem of classification of multiple visual observations has started to receive more and more attention. These multiple observations are typically collected by a network of vision sensors, where each sensor observes the same signal from different viewpoints or at different time instants or even under different noise conditions. The problem in such settings consists in finding the correct class of the underlying signal from a set of candidate classes.

Dimensionality reduction techniques are widely used for solving classification problems. It has been shown that techniques like PCA, LDA, ICA [1] lead to a discriminative representation of the signal based on its underlying structure. The new low dimensional representation emphasizes the most meaningful information in the signal and leads to improved classification accuracy. In sensor networks, one may however not have access to the complete signals but only to compressed versions. For example, the signals can be given under the form of random projections or compressed measurements. Such measurements have been shown to preserve the structure of the original signal with reduced sensing costs [2], so that they can be used for classification instead of the original signals. Furthermore, when multiple compressed versions of the same signal are available, the

amount of information about the signal also increases which usually leads to better analysis or classification.

In this work, we aim at extending the dimensionality reduction properties of random projections by combining them with classical dimensionality reduction techniques in a multiple observations framework. In particular, we focus on the design of a smart fusion center that reaches a global classification decision by aggregating multiple noisy measurements of the same signal while avoiding the expensive cost of signal reconstruction. We extract more information about the signal by considering that the observations are captured with different sensing matrices. Our framework is based on embedding the low dimensional space obtained by linear dimensionality reduction techniques in the compressed domain. The coefficients of the multiple measurements in the embedded subspace are used to classify the signal with a Nearest Neighbor (NN) classifier. We illustrate the performance of the proposed scheme in image classification and we confirm that linear dimensionality reduction techniques increase the classification accuracy even when the signals are given with compressed measurements.

The use of random projections for dimensionality reduction has received considerable interest over the past few years. The authors in [3] compare random projection and PCA using different machine learning methods and confirm that random projections are extremely attractive mainly due to their computational advantages. Moreover, random projections have been applied on various types of compressive signal processing problems [4] including compressive classification [5–7] and face recognition [8]. However, these works usually assume that the original signal space guarantees good classification, which is not always the case. The combination of both linear dimensionality reduction (which leads to more discriminant features) and compressed measurements (which provide simple and effective sensing) is expected to offer improved classification performance with a reduced number of measurements.

## 2. COMPRESSED CLASSIFICATION FRAMEWORK

We examine the problem of classification of multiple compressed observations obtained under different noisy conditions in a sensor network as illustrated in Fig.1. We consider a network that consists of  $S$  generic sensors, each of which obtains linear measurements of a noisy version  $x_n = x + \epsilon_n$  of the signal  $x \in \mathbb{R}^N$ . For a sensor  $n \in \{1, \dots, S\}$ , the vector of measurements is denoted as  $y_n = \Phi_n(x + \epsilon_n)$ , where  $\Phi_n$  is a  $M \times N$  ( $M < N$ ) matrix representing its sampling system. The entries of the matrix are independent and identically distributed (i.i.d.) random variables from a normal

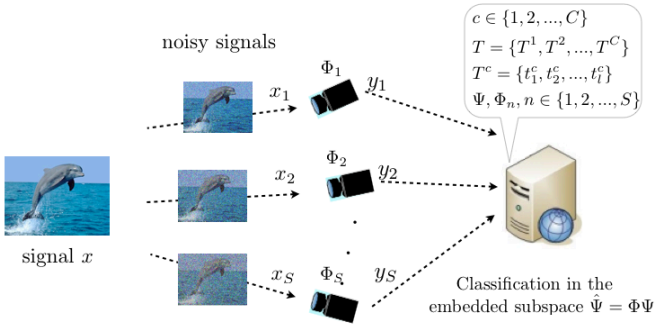


Fig. 1. Sensor network architecture

distribution  $\mathcal{N}(0, 1/M)$ . The noise  $\epsilon_n$  consists of i.i.d., zero-mean, Gaussian random variable with variance  $\sigma^2$ . Suppose that the test signal  $x$  belongs to  $c$ , one of the  $C$  classes, where  $c \in \{1, 2, \dots, C\}$ . We denote our set of training samples as  $T = \{T^1, T^2, \dots, T^C\}$ , with  $T^c = \{t_1^c, t_2^c, \dots, t_f^c\}$  where  $t_i^c$  is the  $i^{\text{th}}$  training vector in class  $c$ .

Our goal is to exploit the diversity achieved by the multiple observations and the use of multiple measurement matrices  $\Phi_n$  in order to reach a classification decision. We assume that there exists a  $K$ -dimensional linear subspace  $\Psi$  where the signal can be classified efficiently with simple nearest neighbor search among the projection coefficient vectors. This classification subspace is independent of the sensing system and can be computed by classical linear dimensionality reduction techniques applied on the training data. We however cannot classify directly the observation sets in  $\Psi$  since we only have access to compressed measurements and not to the observed signals. In addition, we want to avoid the costly reconstruction of the signal because it will introduce artifacts as the sensing matrix  $\Phi_n$  is not full rank in general. Hence, the role of the central node is to reach a classification decision based on the measurements from multiple observations while properly exploiting the knowledge of the classification subspace. The compressed classification problem can be formally defined as follows.

**Compressed Classification Problem.** Given a linear classification subspace  $\Psi \in \mathbb{R}^N$  and a set of classes  $C$ , and given a set of  $S$  sensors with sensing matrices  $\Phi_n$  of size  $M \times N$  ( $M < N$ ), the fusion center has to predict the correct class  $c \in C$  of an object of interest  $x$  from a set of  $S$  compressed noisy observations  $y_n = \Phi_n(x + \epsilon_n)$ .

### 3. CLASSIFICATION OF COMPRESSED OBSERVATION SETS

In this section, we introduce classical dimensionality reduction techniques for solving the problem of classification of multiple compressed observations. We define a new classification subspace by embedding the subspace  $\Psi$  in the compressed domain. The embedding of  $\Psi$  in  $\mathbb{R}^M$  is simply performed by projecting its vectors onto the subspace defined by the rows of the measurement matrix  $\Phi_n$ . The new subspace is characterized by the vectors of the matrix:

$$\hat{\Psi}_n = \Phi_n \Psi. \quad (1)$$

Notice that the classification subspace  $\hat{\Psi}_n$  could also be defined by applying linear dimensionality reduction directly on the set of training measurements. That, however, would increase significantly the computational cost since it would lead to a different subspace  $\hat{\Psi}_n$  for each sensing matrix  $\Phi_n$ .

The coefficients of the projection of the measurements  $y_n$  and the training data onto the subspace defined by the vectors  $\hat{\Psi}_n$  are respectively

$$\hat{\alpha}_{y_n} = \hat{\Psi}_n^T \Phi_n(x + \epsilon_n) = (\Phi_n \Psi)^T \Phi_n(x + \epsilon_n),$$

$$\hat{\alpha}_{t_i^c} = \hat{\Psi}_n \Phi_n t_i^c = (\Phi_n \Psi)^T \Phi_n t_i^c.$$

We exploit the diverse observations that belong to the same class and extract more information about the signal  $x$  by taking their average:

$$\hat{\alpha}_{y_{avg}} = \frac{1}{S} \sum_{n=1}^S \Psi^T \Phi_n^T \Phi_n(x + \epsilon_n). \quad (2)$$

The averaging scheme is expected to work well since we assume that the noisy terms are independent and identically distributed. Thus, the error components with different signs will cancel out in the summation. The single test sample that occurs after averaging will be classified by using the NN classifier:

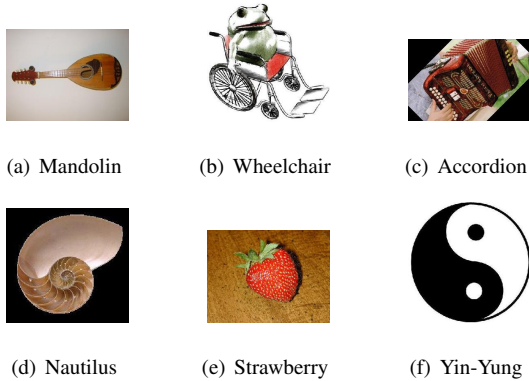
$$c = \underset{c \in \{1, \dots, C\}, \forall i}{\operatorname{argmin}} \|\hat{\alpha}_{y_{avg}} - \frac{1}{S} \sum_{n=1}^S \Psi^T \Phi_n^T \Phi_n t_i^c\|_2^2. \quad (3)$$

We show now that the matrix  $\frac{1}{S} \sum_{n=1}^S \Phi_n^T \Phi_n$  will tend to the identity matrix as we increase the size of the observation sets. Consider one term  $n$  of the summation and denote as  $\phi_{ij}^{(n)}$  the element of the  $i^{\text{th}}$  row and the  $j^{\text{th}}$  column of the matrix  $\Phi_n$ . Let  $\{Q^{(n)}\}$  (with  $n = 1..S$ ) be defined as  $Q^{(n)} = \Phi_n^T \Phi_n$  and denote by  $q_{ij}^{(n)}$  the entries of the matrix  $Q^{(n)}$ . Since the entries of the matrix  $\Phi_n$  are i.i.d. with  $\sigma^2 = 1/M$  we have

$$\begin{aligned} \mathbf{E}\{q_{ii}^{(n)}\} &= \mathbf{E}\left\{\sum_{k=1}^M \phi_{ki}^{(n)2}\right\} = \sum_{k=1}^M \mathbf{E}\{\phi_{ki}^{(n)2}\} = M\sigma^2 = 1 \\ \mathbf{Var}\{q_{ii}^{(n)}\} &= \mathbf{E}\{(q_{ii}^{(n)})^2\} - \mathbf{E}\{q_{ii}^{(n)}\}^2 = \mathbf{E}\left\{\left(\sum_{k=1}^M \phi_{ki}^{(n)2}\right)^2\right\} - 1 \\ &= \mathbf{E}\left\{\sum_{k=1}^M \phi_{ki}^{(n)4} + 2 \sum_{l>k} \phi_{ki}^{(n)2} \phi_{li}^{(n)2}\right\} - 1 \\ &= \sum_{k=1}^M \mathbf{E}\{\phi_{ki}^{(n)4}\} + 2 \sum_{k=1}^M \sum_{l>k} \mathbf{E}\{\phi_{ki}^{(n)2} \phi_{li}^{(n)2}\} - 1 \\ &= M \frac{3}{M^2} + \frac{2}{M^2} \sum_{r=1}^{M-1} (M-r) - 1 = \frac{2}{M}, \end{aligned}$$

where we used the fact that  $\mathbf{E}\{\phi_{ki}^{(n)4}\} = 3 \cdot \sigma^4 = 3/M^2$  and  $\mathbf{E}\{\phi_{ki}^{(n)2} \phi_{lj}^{(n)2}\} = \sigma^2 \cdot \sigma^2 = 1/M^2$ . From the central limit theorem, as the sample size  $S$  increases, the distribution of the sample average  $\overline{\phi_{ii}} = \frac{1}{S} \sum_{n=1}^S q_{ii}^{(n)}$  of these random variables approaches the normal distribution with mean  $\mathbf{E}\{q_{ii}^{(n)}\}=1$  and variance  $\mathbf{Var}\{q_{ii}^{(n)}\}/S = 2/M^2$ . Similarly, for the non diagonal elements of the matrix  $Q^{(n)}$  we compute that  $\mathbf{E}\{q_{ij}^{(n)}\}=0$  and  $\mathbf{Var}\{q_{ij}^{(n)}\} = \frac{1}{M}$ , for  $i \neq j$ . The distribution of the sample average  $\overline{\phi_{ij}} = \frac{1}{S} \sum_{n=1}^S q_{ij}^{(n)}$  thus approaches also a normal distribution with  $\mathcal{N}(0, \frac{1}{MS})$  for large values of  $S$ .

The above equations imply that, as the number of the observation set increases, the variance of the variables  $\overline{\phi_{ii}}$  and  $\overline{\phi_{ij}}$  decreases. The variables tend to their expected value  $\overline{\phi_{ii}} \rightarrow 1$  and  $\overline{\phi_{ij}} \rightarrow 0$ , which means that  $\overline{\Phi} = \frac{1}{S} \sum_{n=1}^S \Phi_n^T \Phi_n \rightarrow I$  for  $S$  sufficiently large. Moreover, for the same variance, the number of measurements that



**Fig. 2.** Example images from each class

are necessary for classification depends on the number of the observations. More precisely, as we increase the size of the observation set, the necessary measurement rate is smaller. The above result confirms that  $\Psi^T \frac{1}{S} \sum_{n=1}^S \Phi_n^T \Phi_n x \rightarrow \Psi^T x$  for large values of  $S, M$ . Hence, the classification performance will get closer to the one obtained by classifying directly the original signal  $x$  in the subspace defined by  $\Psi$ .

## 4. SIMULATION RESULTS

### 4.1. Experimental setup

In this section we quantify the performance of the proposed framework. We use two data sets from the Caltech database<sup>1</sup> for our experimental evaluation, the first containing images of objects belonging to different categories and the second consisting of 450 face images of 27 people. For the first data set, we have selected six categories that will be the different classes of our problem. Each category contains 50 to 100 images and an example image for each class is shown in Fig.2. The classification subspace is found by performing Principal Component Analysis (PCA) to our training data [1]. All the data are then projected onto the subspace defined by the  $K = 10$  largest eigenvectors that are given by PCA. For the second set of experiments, we have selected randomly 10 different individuals from the face data set, under different lighting conditions or with different expressions or backgrounds. The classification subspace is found by applying Linear Discriminant Analysis (LDA) [1] on the training data and the reduced space consists of  $K = 9$  eigenvectors. For computational convenience, all images in both data sets are converted to greyscale and are downsampled to a resolution of 50x50 pixels. For each class, half of the images have been randomly selected as training and the rest as test data.

### 4.2. Classification performance

We explore the effect of the number of the measurements  $M$  on the performance of the NN classifier for both data sets. We first consider noisy samples of the signal  $x_n = x + \epsilon_n$ . The noise  $\epsilon_n$  is modeled as zero mean, additive Gaussian noise with  $\sigma = 0.8$  which corresponds to an SNR between 14 and 20 dB, depending on the energy of each test image. We compare the classification performance of the NN classifier in four different classification schemes: (i) classification of the signal  $x_{avg} = \frac{1}{S} \sum_{n=1}^S x_n$  in  $\mathfrak{R}^N$ , (ii) classification of the signal

$x_{avg}$  in the classification subspace  $\Psi$ , (iii) classification of the measurements  $y_{avg} = \frac{1}{S} \sum_{n=1}^S \Phi_n x_n$  in  $\mathfrak{R}^N$  and (iv) proposed framework. We perform classification experiments for different numbers of compressive measurements,  $M = [10 : 10 : 1000]$  (in MATLAB notation). For illustration, we select the size of the observation set as  $S = 20$ . We measure the performance in terms of classification error rate, which is the percentage of the test samples that have been misclassified. For each setting we execute 100 random experiments of observation sets, which correspond to 100 different realizations of the measurement matrix  $\Phi_n$  and the noise term  $\epsilon_n$ . For each value of  $M$ , we report the average classification error rate.

Fig.3 verifies that, for both data sets, linear dimensionality reduction techniques lead to a better separation of the data that belong to the different classes, resulting in lower classification error rates. In particular, the classification of signals in the subspace  $\Psi$  is more accurate in comparison to the classification performance in the high dimensional space  $\mathfrak{R}^N$ . In the case of compressed measurements, we observe that the embedding of the classification subspace  $\Psi$  in the compressed domain gives better classification results than the direct classification of the measurements in  $\mathfrak{R}^N$ . Notice that the framework works for both PCA and LDA and it does not depend on the linear dimensionality reduction technique that is applied to the training data. Moreover, our classification framework approximates the performance of the classification of the signal in  $\Psi$  with only a few measurements.

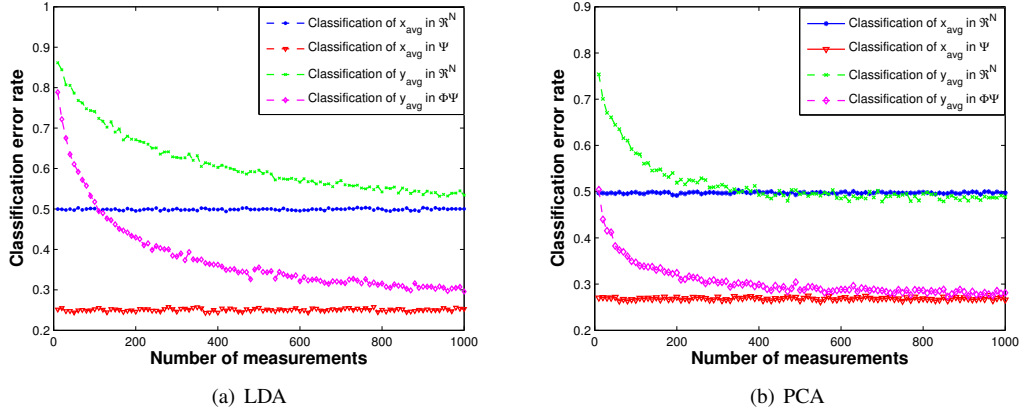
In addition, we examine the performance of the averaging scheme for different sizes of observation sets. We illustrate the classification performance in the cases of (i) a single measurement matrix, i.e.,  $\Phi_n = \Phi, \forall n$  and (ii) multiple measurement matrices  $\Phi_n$ . We observe in Fig.4 that an increase in the number of observations gradually improves the classification error rate in both cases. Notice, first, that the measurement rate for a specific error rate reduces as the size of the observation set increases. Moreover, the classification accuracy improves and in particular, for a sufficiently large value of  $S$ , the classification performance of the compressed measurements will tend to the one obtained using high dimensional signals. This improvement is expected because as we have already shown in Section 3, the larger the number of observations, the higher the probability that the matrix  $\bar{\Phi} = \frac{1}{S} \sum_{n=1}^S \Phi_n^T \Phi_n$  tends to the identity matrix  $I$ .

Finally, we compare the performance in the cases where sensors have a single or respectively multiple measurement matrices. It is obvious from Fig.4 that the performance is significantly improved in the latter case. Since each  $\Phi_n$  is different, all their rows effectively combine to span more space in  $\mathfrak{R}^N$ . Eq.(2) results in a linear combination of multiple transformed representations of the same signal, meaning that the performance of the classifier in the compressed domain will approximate the performance in the subspace  $\Psi$ . On the other hand, if all the observations are obtained with the same matrix  $\Phi$  then all the measurements will belong to the same subspace which only covers part of  $\mathfrak{R}^N$ . Moreover, the central limit theorem and thus our analysis in Section 3 do not hold in this case since the variables  $\{q_{ii}^{(n)}\}, n = 1..S$  are not independent. For an increasing  $S$ , the performance is still expected to improve, due to the fact that the increase in the size of the observation set will effect the error terms. However, this improvement will be smaller in comparison to the achievable gain in the multiple measurement matrices case.

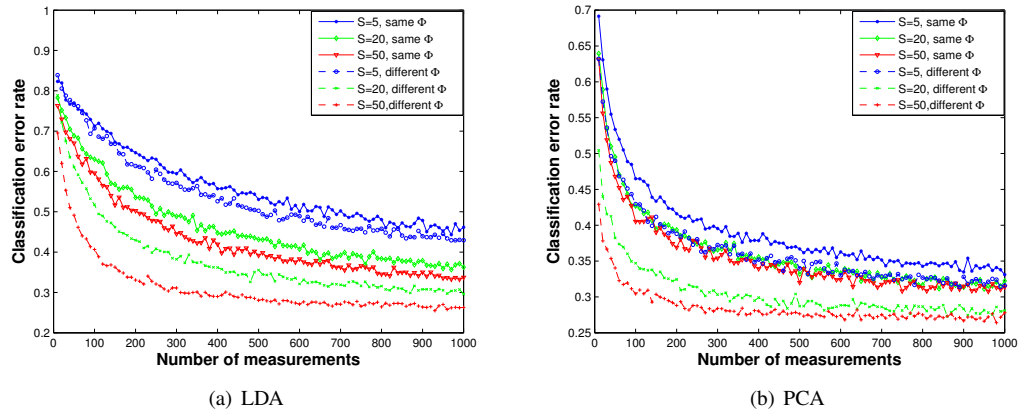
## 5. CONCLUSIONS

In this paper we studied the problem of classification of multiple compressed observations with linear subspace embeddings. We

<sup>1</sup>Publically available at: <http://www.vision.caltech.edu/>



**Fig. 3.** Classification performance of (i) the noisy signals in  $\mathcal{R}^N$ , (ii) the noisy signals in  $\Psi$ , (iii) the compressed measurements in  $\mathcal{R}^N$  and (iv) the proposed framework on the Caltech data sets of (a) faces and (b) objects.



**Fig. 4.** Classification performance for different sizes of observation sets on the Caltech data sets of (a) faces and (b) objects.

introduced the use of classical linear dimensionality reduction techniques and multiple observations of the same signal in order to reduce the number of low cost random projections needed to correctly classify a signal. The proposed classification framework maintains the good classification properties of the dimensionality reduction techniques by embedding the classification subspace in the compressed domain. Experimental results show that the proposed scheme generally performs well, while only a few measurements are sufficient to achieve high classification accuracy.

## 6. REFERENCES

- [1] C. M. Bishop, *Pattern recognition and machine learning*, Springer, 2006.
- [2] D. Achlioptas, "Database-friendly random projections," in *Proc. of the 20th ACM symposium on Principles of Database Systems*, Santa Barbara, CA, USA, 2001, pp. 274–281.
- [3] D. Fradkin and D. Madigan, "Experiments with random projections for machine learning," in *Proc. of the 9th ACM SIGKDD international conference on Knowledge discovery and data mining*, Washington, DC, USA, 2003, pp. 517–522.
- [4] A. Davenport, T. Boufounos, B. Wakin, and G. Baraniuk, "Signal processing with compressive measurements," *IEEE Journal of Selected Topics in Signal Processing*, vol. 4, no. 1, pp. 445–460, Feb. 2010.
- [5] M. Davenport, M. Duarte, M. Wakin, J. Laska, D. Takhar, K. Kelly, and R. Baraniuk, "The smashed filter for compressive classification and target recognition," in *Proc. IS&T/SPIE Symposium on Electronic Imaging: Computational Imaging*, San Jose, CA, USA, Jan. 2007.
- [6] M. Davenport, M. Duarte, M. Wakin, J. Laska, D. Takhar, K. Kelly, and R. Baraniuk, "Multiscale random projections for compressive classification," in *Proc. IEEE Int. Conference on Image Processing*, San Antonio, TX, USA, Sept. 2007.
- [7] J. Haupt, R. Castro, and R. Nowak, "Compressive sampling for signal classification," in *Proc. of 40th Asilomar Conf. Signals, Systems and Computers*, Pacific Grove, CA, USA, Nov. 2006.
- [8] N. Goel, G. Bebis, and A. Nefian, "Face recognition experiments with random projection," in *SPIE Conference on Biometric Technology for Human Identification*, Orlando, FL, USA, Apr. 2005.