SPARSE REGRESSION ANALYSIS FOR OBJECT RECOGNITION

Baochang Zhang[1], *Shengping Zhang*[2], *Jianzhuang Liu*[3,4]

[1]School of Automation Science and Electrical Engineering, Beihang University, China
[2]School of Computer Science and Technology, Harbin Institute of Technology, China
[3]Department of Information Engineering, The Chinese University of Hong Kong, China
[4]Shenzhen Institutes of Advanced Technology, Chinese Academy of Sciences, China

ABSTRACT

This paper proposes a new method named Sparse Regression Analysis (SRA) for object representation and recognition. In SRA, ℓ_1-norm minimization is combined with regression analysis to represent the input signal. The discriminative ability of SRA derives from the fact that the subset which most compactly expresses the input signal is activated in the regression analysis. To achieve a further improvement, Kernelized SRA (KSRA) is developed to make a nonlinear extension of SRA. The experiments are conducted on both palmprint and face recognition, which show that the proposed methods achieve a much better performance than sparse representation classifier, principal component analysis, and linear discriminant analysis.

***Index Terms*—** Sparse representation, ℓ_1 norm minimization, face recognition, palmprint recognition

1. INTRODUCTION

Object representation in parsimony is one of the important principles for object recognition. One initial work is the principle of minimum description length in model selection [1], which yields the most compact representation for decision-making tasks such as classification. The Small Sample Size (SSS) statistic learning acquires a parsimony representation based on only a few of the observations by selecting a small subset of features for classification or visualization. Support vector machine (SVM) [2] is one of the most successful methods using a small subset of relevant training examples to characterize the decision function between classes. Considering that many signals and images contain redundant information, feature reduction aiming to represent data in a parsimony way is often an important step for a pattern recognition system. One popular approach for compressing the input signal is the Principal Component Analysis (PCA) [3]. It transforms a number of correlated variables into a small number of uncorrelated variables called principal components, by exploiting the fact that many signals have a sparse representation in terms of some basis. The discriminant versions of component analysis have been thoroughly investigated, such as Linear Discriminant Analysis (LDA) [4]. One problem lies in these methods is that they are optimal when the data are in Gaussian distribution, and the generalization problem is not well solved yet. These works mentioned above indicate a common theme: using parsimony as the principle for choosing a limited subset of features or models from the training data, rather than directly using the data for representing or classifying the input signal or object, is more effective for object representation and recognition.

Investigation and study in the human vision system (HVS) have shown that a selective and small subset of neurons is active for a variety of specific stimuli, such as color, texture, shape, and scale. Considering a large amount of neurons in the human vision system, the firing of the neurons to a specific input object is typically highly sparse. In the statistical signal processing community, the algorithmic problem of computing sparse linear representations of an over-complete dictionary of basis components or basis samples has been well developed. In [5], a promising application of sparse representation in building a classifier for face recognition shows that the sparse representation is effective for classification even using a simple ℓ_1 -norm minimization.

In this paper, we propose a new method, named Sparse Regression Analysis (SRA), for discriminant analysis on object recognition. Intuition behind SRA lies in the fact that the coefficients of the subset compactly expressing the input object are nonzero, and others are almost zero, when combining ℓ_1-norm minimization and regression analysis for the reconstruction of the input object. It is possible to automatically represent a test object as a linear combination of only those training samples from the same class or similar samples. The coefficients of the sparse representation therefore automatically discriminate different classes based on a common basis. The proposed method is different from other related methods [5] in three aspects. First, our method focuses on feature extraction, but other others focus on reconstruction error or its application to build a classifier [5] in object recognition. Second, we extend the sparse representation to the kernelized version to obtain nonlinear features. Last, the proposed method can be used to extract discriminant features

978-1-4577-1302-6/11/$26.00 ©2011 IEEE

from a training set, but without class information (i.e. the class labels of the training samples), which has been used as in traditional discriminant methods, such as LDA and KDA.

2. SPARSE REGRESSION ANALYSIS

2.1. Sparse Representation

When we say an object $\boldsymbol{y}$ has a sparse representation, we mean that $\boldsymbol{y}$ is well approximated by a linear combination of a small subset of a vector set $\boldsymbol{\Psi}$ of size N , i.e., $\sum_{i\in\mathbb{I}_K} w_i\boldsymbol{\psi}_i$, where $\mathbb{I}_K$ is a subset of $\{1,2,\ldots,N\}$ with K being an integer, $K \ll N$, and w_i is the weight corresponding to $\boldsymbol{\psi}_i \in \Psi$. In this case, we say that $\boldsymbol{y}$ is K-sparse in $\boldsymbol{\Psi}$. The number of the non-zero coefficients is denoted by $\|\boldsymbol{w}\|_0$ with $\boldsymbol{w} = (w_1, w_2, \ldots, w_N)^T$. Minimizing $\|\boldsymbol{w}\|_0$ is the principle to obtain a sparse representation, which is, however, an NP-hard problem. Recent development in the theory of sparse representation shows that the solution of ℓ_1-norm minimization subject to the linear regression of the input sample can be used to find sparse enough representation. The resulting optimization problem, similar to the LASSO in statistics [6], penalizes the ℓ_1-norm of the coefficients in the linear combination, rather than directly penalizing the number of nonzero coefficients ($\|\boldsymbol{w}\|_0$). The original works in this research are not for classification, but for representation of signals. The algorithm performance is measured in terms of sparsity of the representation and reconstruction of the signals. It is possible to represent a test sample as a linear combination of only those training samples from the same class. The coefficients of the sparse representation therefore automatically discriminate different classes present in a given basis on a dictionary of basis elements. In [5], Wright et al. provide a general classifier using the sparse representation in the face recognition problem. Its validity has been testified extensively on public databases.

2.2. Sparse Regression Analysis

In this section, we exploit the discriminative ability of sparse regression representation to perform feature extraction. We represent a test sample in a dictionary whose basis elements are the training samples themselves, when sufficient training samples are available. To recover a signal from a training set, we choose to minimize $\|\boldsymbol{w}\|_1$. In this model, all test samples are constructed from the same set of the basis vectors (the training samples), with different coefficients. The problem is formulated as follows:

$$\text{minimize:} \quad \|\boldsymbol{w}\|_1, \tag{1}$$

$$\text{subject to} \quad \Psi\boldsymbol{w} = \boldsymbol{y}, \tag{2}$$

where $\Psi = (\boldsymbol{x}_1, \boldsymbol{x}_2, \ldots, \boldsymbol{x}_N)$ is a matrix with each training sample arranged on its columns, and Eq. 2 is the regression representation of the input signal $\boldsymbol{y}$, which can be evaluated by $\|\Psi\boldsymbol{w} - \boldsymbol{y}\|_2 = 0$. The model shown in Eq. 1 can be generalized to:

$$\text{minmize} \quad \|\Psi\boldsymbol{w} - \boldsymbol{y}\|_2 + \lambda\|\boldsymbol{w}\|_1, \tag{3}$$

For some practical applications, a linear model cannot obtain good performance. To solve nonlinear problems, we exploit the kernel trick to make a nonlinear extension of the original model. In kernel-based methods, the input data is first projected into an implicit feature space $\boldsymbol{F}$ by a nonlinear mapping $\Phi : \boldsymbol{x} \rightarrow \boldsymbol{f} \in \boldsymbol{F}$. In a high-dimensioned space, we define a new constraint function as:

$$\Phi(\Psi)\boldsymbol{w} = \Phi(\boldsymbol{y}), \tag{4}$$

where $\Phi(\Psi) = (\Phi(\boldsymbol{x}_1), \Phi(\boldsymbol{x}_2), \ldots, \Phi(\boldsymbol{x}_N))$, and Φ is the projection function mapping signals from the original space to the high-dimensional space. We cannot directly calculate Eq. 4 because the mapping function is implicit. To implement it, we rewrite Eq. 4 as

$$\Phi^T(\Psi)\Phi(\Psi)\boldsymbol{w} = \Phi^T(\Psi)\Phi(\boldsymbol{y}). \tag{5}$$

The inner product in the high-dimensional space can be calculated by using a kernel function $k(\cdot,\cdot)$ defined in the original space as

$$\Phi(\boldsymbol{x}_i) \cdot \Phi(\boldsymbol{x}_j) = k(\boldsymbol{x}_i, \boldsymbol{x}_j). \tag{6}$$

We use the fractional kernel function [7] in our application, which is defined as:

$$k(\boldsymbol{x}_i, \boldsymbol{x}_j) = (\boldsymbol{x}_i^T \cdot \boldsymbol{x}_j)^{\frac{1}{n}}, \quad n \geq 1, \text{ is an interger.} \tag{7}$$

The evidence that similar images of the same class can well recover the input signal [5] motivates us to use the coefficients $w_i, 1 \leq i \leq N$ for extracting features to discriminate among different classes, which provides a new approach for object classification. In contrast, statistic methods such as linear discriminant analysis that utilizes training samples generally incorporate more explicit prior knowledge about the types of sample variations.

2.3. Classification with the Sparse Regression Representation

Our use of the sparse representation for classification differs from the other related feature representation techniques. Instead of using the sparsity to identify a relevant model or classifier that can later be used for classifying test samples, we use the sparse regression coefficients of each individual test sample directly for classification, adaptively selecting the training samples that give the most compact representation. The proposed feature extraction can be easily combined with popular classifiers such as nearest neighbor (NN). For the object recognition problems discussed in the next sections we simply use the NN classifier with the cosine distance being the

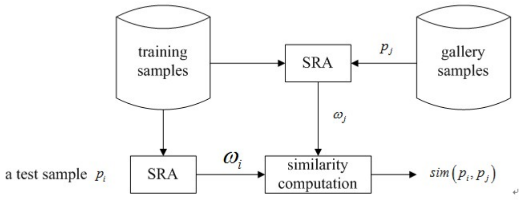


Fig. 1. Computation of the similarity between two samples.

similarity measure, which classifies a test sample based on its best match image in the gallery set.

Let $\boldsymbol{w}_i$ and $\boldsymbol{w}_j$ be the features extracted from two samples p_i and p_j by the linear SRA. The similarity between them is computed by:

$$\text{Sim}(p_i, p_j) = \frac{\boldsymbol{w}_i^T \boldsymbol{w}_j^T}{\|\boldsymbol{w}_i\| \|\boldsymbol{w}_j\|}. \tag{8}$$

Fig. 1 shows the flowchart of the computation of the similarity between two samples. Note that all the representation coefficients of the samples in the gallery can be pre-computed offline.

It should be emphasized that before classifying a test sample, it is not needed to include samples of the same class as the test sample in the training set, which however is necessary in the Sparse Representation Classifier (SRC) in [5]. The proposed method is generally based on the similarity property of samples of the same category. For example, it is human faces in face recognition, and it is palmprints in palmprint recognition. To validate the proposed method, we apply it to palmprint and face recognition. In palmprint recognition, we compare our method with Palmcode [8], which is the state-of-the-art. In face recognition, some well-known approaches, Eigenface, Fisherface, and SRC are compared on both the AR and Infrared face databases.

3. EXPERIMENTS

3.1. Palmprint Recognition

We conduct experiments on the Polyu palmprint database [8] which contains 600 images for 100 subjects with 6 images for each person. The protocol includes a training set, a gallery set, and a probe set (containing test images). The first image of each class is included in the gallery set, and the rest are divided into the train and probe sets. Since SRC [5] does not use a gallery set, the training set for it is the combination of the training set and the gallery set. Both SRC and SRA use the same probe set. Compared with SRC, SRA achieves a much better performance when the training set is smaller as shown in Fig. 2. The curve of the recognition rate of SRA remains flatter, while SRC suffers a lot from the reducing size of the training set.

We further compare SRA with the kernelized SRA (KSRA) when $n = 4$ (see Eq. 7). As shown in Fig. 3(a),

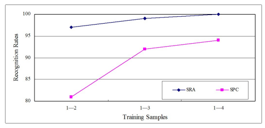
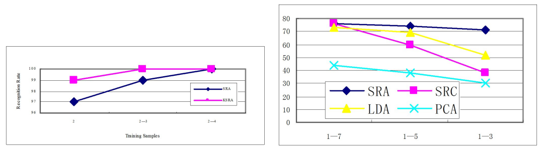


Fig. 2. The comparative results between SRA and SRC, where means samples with indices form 1 to in every class (person) are used for training in SRC.



(a) (b)

Fig. 3. (a)The comparative results between SRA and KSRA. (b)The comparative results among SRA, SRC, LDA, and PCA.

KSRA achieves 100% accuracy in the training set. The performance of KSRA is also evaluated with the sate-of-the-art method Palmcode on this dataset. The Equal Error Rate (EER) of our result is 0% when samples with indices 2 and 3 are in the training set, While the EER of Palmcode is 0.34%.

3.2. Face recognition on the AR Face Database

We test the proposed method on the AR database, which consists of 126 persons. For each person, 26 pictures are taken in two separate sessions [9]. The face regions are cropped, normalized to the 88×88, and converted to grayscale. Similar to [5], we downsize the 16×16-sized face images into 4×8 size subregions so that the feature extraction process does not take too much time. These images include facial variations such as illumination changes, expressions, and facial disguises. For each person, the images with only illumination changes and expressions are selected: the images from Session 1 for training, and the images from Session 2 for testing. We try different numbers of training samples. The gallery set contains all the 14^{th} images in Session 2 in SRA, and the probe set includes images with indices from 15 to 20 from Session 2 in both SRC and SRA. Besides, all the 14^{th} images are included in the training set in SRC.

Fig. 3(b) shows the comparative results among SRA, SRC, LDA and PCA. The results of SRC are 75.3% , 59.52%, and 0.383%, but our results remain stable, which are 76.1%, 73.9%, and 71.55%, when the train set contains 1–7, 1–5, and 1–3 images respectively. The proposed method achieves a much better performance than SRC when using a small set of training samples. The main reason is that SRC uses only part of the training samples to do classification, ignoring other useful information from different classes. The LDA method also suffers a lot from the reducing size of the training set,

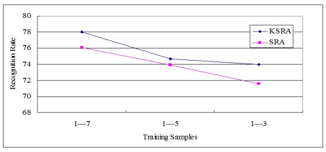


Fig. 4. The comparative results between KSRA and SRA.

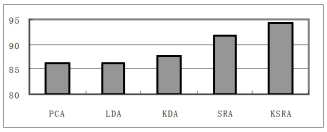


Fig. 5. Experiment results on the infrared face database

because the class discriminative information is lost in training the classification model. However, we can see that LDA obtains a better performance than SRC when a small training set of samples are available. We also give the comparative results between KSRA and SRA in Fig. 4, from which, we can see that KSRA performs even better.

It should be noted that our implementation of SRC follows that in [5]. When seven images from Session 1 for each subject (total 100 subjects are selected) are used for training and the other seven from Session 2 are used for testing, about 85% recognition rate is obtained in [5], while our implementation of their method achieves a similar results of 86%.

3.3. Face Recognition on the PolyU-NIRFD Face Database

Usually infrared face images do not contain as detailed information as grey-level images. The sparse representation can work well on images even with higher degree of downsampling [5]. We choose the PolyU-NIRFD face database [10] in this experiment. The training set contains 419 frontal images of 138 persons, and the probe and gallery sets have 2763 and 574 images, respectively. No persons in the query and target sets appear in the training set. The facial portion of each original image is cropped and normalized to 64×64 pixels based on the locations of the eyes. To see how well PCA, LDA, KDA, SRA and KSRA perform on this database, their recognition performances are shown in Fig. 5. As we know from the experiment on the AR database, SRA and KSRA work better on downsampled images than PCA and LDA that even use the images without downsampling. In this experiment, PCA, LDA, and KDA cannot obtain good performance either, but SRA and KSRA still work well or their robustness to low-resolution images.

4. CONCLUSION

This paper has proposed a new method, named Sparse Regression Analysis (SRA), for discriminant analysis on object recognition. SRA combines the ℓ_1-norm minimization and regression analysis to classify samples of multiple classes. SRA and its kernelized version KSRA are validated on both palmprint and face recognition. Compared with a recently developed method SRC, SRA and KSRA obtain much better performance especially when a small set of training samples is available. SRA and KSRA also achieve better performance than the popular methods PCA, LDA, and KDA.

5. ACKNOWLEDGMENTS

This work was supported in part by the Natural Science Foundation of China, under Contracts 60903065 and 61039003, in part by the Ph.D. Programs Foundation of Ministry of Education of China, under Grant 20091102120001, in part by the Fundamental Research Funds for the Central Universities, and in part by the Key Laboratory of Robotics and Intelligent System, Guangdong Province (2009A060800016).

References

[1] J. Rissanen, "Modeling by shortest data description," *Automatica*, pp. 465–471, 1978.

[2] V. Vapnik, "The nature of statistical learning theory," *Springer*, 2000.

[3] M. Turk and A. Pentland, "Eigenfaces for recognition," *Journal of Cognitive Neuroscience*, vol. 3, pp. 71–86, 1991.

[4] P. Belhumeur, J. Hespanda, and D. Kriegman, "Eigenfaces vs. fisher-faces: recognition using class specific linear projection," *IEEE Trans. Patt. Anal. Mach. Intell.*, vol. 19, no. 7, pp. 711–720, 1997.

[5] J. Wright, A.Y. Yang, A. Ganesh, S.S. Sastry, and Y. Ma, "Robust face recognition via sparse representation," *IEEE Trans. Patt. Anal. Mach. Intell.*, vol. 31, no. 2, pp. 210–227, 2008.

[6] R. Tibshirani, "Regression shrinkage and selection via the lasso," *Journal of the Royal Statistical Society B*, vol. 58, no. 1, pp. 267–288, 1996.

[7] C. Liu, "Capitalize on dimensionality increasing techniques for improving face recognition grand challenge performance," *IEEE Trans. Patt. Anal. Mach. Intell.*, vol. 28, no. 5, pp. 725–737, 2006.

[8] D. Zhang, A.W. Kong, J. You, and M. Wong, "Online palmprint identification," *IEEE Trans. Patt. Anal. Mach. Intell.*, vol. 25, no. 9, pp. 1041–1050, 2003.

[9] A. Martinez, "Recognizing imprecisely localized, partially occluded, and expression variant faces from a single sample per class," *IEEE Trans. Patt. Anal. Mach. Intell.*, vol. 24, no. 6, pp. 748–763, 2002.

[10] B. Zhang, L. Zhang, D. Zhang, L. Shen, and Z.H. Guo, "Directional binary code with application to polyu near-infrared face database," *Pattern Recognition Letters*, vol. 31, no. 14, 2010.