Optik 127 (2016) 3605–3613

Contents lists available at ScienceDirect

Optik

journal homepage: www.elsevier.de/ijleo

Sparsity embedding projections for sparse representation-based classification

Haishun Du*, Qingpu Hu, Manman Jiang, Fan Zhang

Institute of Image Processing and Pattern Recognition, Henan University, Kaifeng 475004, China

ARTICLE INFO

Article history:
Received 27 September 2015
Accepted 29 December 2015

Keywords:
Sparsity embedding projections
Sparse representation-based classification
Feature extraction
Image recognition

ABSTRACT

Sparse representation-based classification (SRC) has become a powerful tool for image recognition. SRC sparsely encodes a test sample over all training samples and then classifies the test sample into the class that generates the minimal reconstruction residual. However, in many real-world applications, nuisances (e.g. illuminations, view directions, pixel corruptions, and occlusion, etc.) may make the representation coefficients of a test sample associated with the training samples from another class greatly larger than those associated with the training samples from the correct class. As a result, the reconstruction residual of the test sample with respect to the other class is smaller than that with respect to the correct class. This inevitably brings a wrong classification of SRC. To address this issue, we propose a sparsity embedding projections (SEP) method, which seeks a low-dimensional embedding subspace where the sparse representation coefficients of a test sample associated with the training samples from the correct class are enlarged, and simultaneously those associated with the training samples from all of the other classes are compressed. Specially, given a training data matrix, SEP tries to find a linear transformation by enhancing the intraclass reconstructive relationship meanwhile suppressing the interclass reconstructive relationship in the low-dimensional embedding subspace. Experimental results on the COIL-20, Extend Yale B, and AR databases show that the proposed method is more effective and robust than other state-of-the-art feature extraction methods with respect to SRC.

1. Introduction

Image recognition (e.g. face recognition) has become a hot topic in areas of computer vision and pattern recognition. Recently, Wright et al. [1] presented a sparse representation-based classification (SRC) method, and successfully applied it to facial image classification. SRC sparsely encodes a test sample over a dictionary consisting of training samples via L_1-norm optimization technique, and then classifies the test sample into the class that generates minimal reconstruction residual. In [1], they claimed that the choices of feature are not critical as long as the sparsity is properly harnessed. However, in many real-world applications, the dimension of data is usually much larger than the number of training samples, which inevitably increases computational burden and makes the obtained solution unstable. Thus, feature extraction is an essential procedure before performing SRC in most of real-world applications. Moreover, some researchers [2,3] argued that when the number of features is relatively small, the performance between different feature extraction methods is significantly different. They also pointed out that a small number of representative and discriminative features can greatly improve the classification performance, which is more suitable for real-world applications [2,3].

In the past few decades, many subspace learning methods, such as principle component analysis (PCA) [4], linear discriminant analysis (LDA) [5], locality preserving projections (LPP) [6], marginal Fisher analysis (MFA) [7], have been proposed for image feature extraction. Although these subspace learning methods have been successfully used to solve many image recognition problems, they may be not robust to nuisances (e.g. illuminations, view directions, pixel corruptions, and occlusion, etc.) of images. Especially, when the number of training samples is small, the learned subspace may be deflective [8]. Recently, sparse representation model has been used in feature extraction. Qiao et al. [9] proposed a novel feature extraction method, namely sparsity preserving projection (SPP), in which they introduced sparse representation model into the general graph embedding framework and constructed a L_1-graph that possesses advantage of the typical k-nearest neighborhood graph and contains natural discriminative information. Specifically, SPP first sparsely encodes each sample over all samples, and then seeks a set of projections that can preserve the sparse

* Corresponding author. Tel.: +86 15938576856.
E-mail address: jddhs@henu.edu.cn (H. Du).

http://dx.doi.org/10.1016/j.ijleo.2015.12.169

reconstructive relationship of the data [9]. However, SPP does not consider the label information of samples, which is essential for image classification. To enhance the discriminability of SPP, Zhang et al. [10] proposed a discriminative learning by sparse representation projections (DLSP) method by combining the merits of both local interclass geometry and sparsity property. Lai et al. [11] presented a global sparse representation projections (GSRP) method, which seeks a low-dimensional subspace by preserving the sparse reconstructive relationship of the data meanwhile maximizing the interclass separability. In [12], Qiao et al. also proposed a semi-supervised feature extraction method, namely sparsity preserving discriminant analysis (SPDA) for single training sample scenario.

To fit SRC well, Yang et al. [2] proposed a SRC steered discriminative projection (SRC-DP) method based on the decision rule of SRC, which seeks a set of projections by maximizing the ratio of the between-class reconstruction residual to the within-class reconstruction residual in the low-dimensional subspace. However, the convergence of SRC-DP remains unclear, and it is also time-consuming due to the computation burden of iterative sparse coding [3]. To improve the computation efficiency of SRC-DP, Lu et al. [3] proposed an optimized projections for sparse representation based classification (OP-SRC) method, which obtains the projections by solving a generalized eigenvalue problem.

In this paper, we propose a novel feature extraction method, namely sparsity embedding projections (SEP) for SRC. Since SRC classifies a test sample based on the class reconstruction residuals, the dominant representation coefficients of the test sample should be concentrated on the correct class. As observed in [1], nuisances may make the representation coefficients of a test sample associated with the training samples from another class greatly larger than those associated with training samples from the correct class. It will make SRC misclassification. To overcome this issue, we try to seek a low-dimensional embedding subspace where the sparse representation coefficients of a test sample associated with the training samples from the correct class are enlarged, and simultaneously those associated with the training samples from all of the other classes are compressed. As a result, the reconstruction residual of the test sample given by the correct class is much smaller than those given by all of the other classes, which undoubtedly brings a correct classification of SRC.

The rest of this paper is organized as follows. Section 2 briefly reviews sparse representation-based classification (SRC). The proposed SEP method is presented in Section 3. Experimental results on several image databases and corresponding discussions are presented in Section 4. Finally, Section 5 concludes this paper.

2. Sparse representation-based classification

The sparse representation-based classification (SRC) was proposed in [1] for robust facial image recognition. Suppose there are C classes of training samples, let $\mathbf{X}_i = [\mathbf{x}_{i1}, \mathbf{x}_{i2}, \ldots, \mathbf{x}_{iNi}] \in \mathrm{R}^{D\times N_i}$ be the matrix formed by N_i training samples of ith class, in which $\mathbf{x}_{ij}$ is an vector stretched by the jth sample of the ith class. Then $\mathbf{X} = [\mathbf{X}_1, \mathbf{X}_2, \ldots, \mathbf{X}_C] \in \mathrm{R}^{D\times N}$ is the matrix of all training samples, where $N = \sum_{i=1}^{C} N_i$ is the total number of training samples. Using $\mathbf{X}$ as dictionary, a test sample $\mathbf{y}$ can be approximately represented as linear combination of the elements of $\mathbf{X}$, i.e. $\mathbf{y} = \mathbf{X}\boldsymbol{\alpha}$, in which $\boldsymbol{\alpha}$ is the representation coefficient vector of $\mathbf{y}$. If $D < N$, the linear equation $\mathbf{y} = \mathbf{X}\boldsymbol{\alpha}$ is underdetermined. Therefore, the solution $\boldsymbol{\alpha}$ is not unique. Since the test sample $\mathbf{y}$ can be adequately represented by the training samples from the correct class, the representation is obviously sparse when the amount of training samples is large enough. SRC claims that the sparser the representation coefficient vector $\boldsymbol{\alpha}$ is, the easier it is to recognize the test sample's class label [1]. This inspires one to find the sparsest solution of $\mathbf{y} = \mathbf{X}\boldsymbol{\alpha}$ by solving the following optimization problem:

$$\hat{\boldsymbol{\alpha}} = arg \min_{\boldsymbol{\alpha}} \left\|\mathbf{y} - \mathbf{X}\boldsymbol{\alpha}\right\|_2 + \lambda\|\boldsymbol{\alpha}\|_1, \tag{1}$$

where $\lambda > 0$ is a scalar constant. The optimization problem (1) can be efficiently solved by many algorithms, such as basis pursuit [13], l1_ls [14] and alternating direction algorithm [15].

Having obtained the sparsest solution $\hat{\boldsymbol{\alpha}}$, let $\delta_i : \mathrm{R}^N \to \mathrm{R}^N$ be the characteristic function that chooses the coefficients associated with the ith class. For $\hat{\boldsymbol{\alpha}} \in \mathrm{R}^N$, $\delta_i\left(\hat{\boldsymbol{\alpha}}\right)$ be a vector, whose only nonzero entries are the entries in $\hat{\boldsymbol{\alpha}}$ that are associated with class i [1]. Using $\delta_i\left(\hat{\boldsymbol{\alpha}}\right)$, the test sample $\mathbf{y}$ can be reconstructed as $\hat{\mathbf{y}}_i = \mathbf{X}\delta_i\left(\hat{\boldsymbol{\alpha}}\right)$. Then, the test sample $\mathbf{y}$ is classified into the class that minimizes the class reconstruction residual between $\mathbf{y}$ and $\hat{\mathbf{y}}_i$:

$$\text{identity}(\mathbf{y}) = arg \min_i \left\|\mathbf{y} - \mathbf{X}\delta_i\left(\hat{\boldsymbol{\alpha}}\right)\right\|_2. \tag{2}$$

In real-world applications, the observations usually are corrupted or occluded. A test sample $\mathbf{y}$ can be rewritten as:

$$\mathbf{y} = \mathbf{y}_0 + \mathbf{e} = \mathbf{X}\boldsymbol{\alpha} + \mathbf{e} = [\mathbf{X}, \mathbf{I}]\begin{bmatrix}\boldsymbol{\alpha}\\ \mathbf{e}\end{bmatrix} = \mathbf{B}\boldsymbol{\omega}, \tag{3}$$

where $\mathbf{B} = [\mathbf{X}, \mathbf{I}] \in \mathbf{R}^{D\times(D+N)}$, and $\mathbf{I} \in \mathrm{R}^{D\times D}$ is the identity matrix. The clean test sample $\mathbf{y}_0$ and the corresponding error $\mathbf{e}$ can be sparsely encoded over the dictionary $\mathbf{X}$ and the occlusion dictionary $\mathbf{I}$. The sparsest solution $\hat{\boldsymbol{\omega}} = \left[\hat{\boldsymbol{\alpha}}^{\mathrm{T}}, \hat{\mathbf{e}}^{\mathrm{T}}\right]^{\mathrm{T}}$ can be obtained by solving an optimization problem similar to problem (1), and then the test sample $\mathbf{y}$ is classified by the following decision rule:

$$\text{identity}(\mathbf{y}) = arg \min_i \left\|\mathbf{y} - \mathbf{X}\delta_i\left(\hat{\boldsymbol{\alpha}}\right) - \hat{\mathbf{e}}\right\|_2. \tag{4}$$

Although SRC claims that the choices of feature are not critical as long as the sparsity is properly harnessed [1], the representative and discriminative features can dramatically improve its performance [2,3].

3. Sparsity embedding projections

3.1. Formulation

Let $\mathbf{X} = [\mathbf{X}_1, \mathbf{X}_2, \ldots, \mathbf{X}_C] \in \mathrm{R}^{D\times N}$ be the training data matrix composed of C classes of training samples, in which $\mathbf{X}_i = \left[\mathbf{x}_{i1}, \mathbf{x}_{i2}, \ldots, \mathbf{x}_{iN_i}\right] \in \mathrm{R}^{D\times N_i}$ be the matrix formed by N_i training samples of ith class. For each training sample $\mathbf{x}_{ij}$, we leave it out from $\mathbf{X}$, and sparsely encode it over the remaining training samples. This brings the following L_1-optimization problem:

$$\min_{\mathbf{s}_{ij}\geq 0} \quad \left\|\mathbf{s}_{ij}\right\|_1 \quad \text{s.t.} \quad \left\|\mathbf{x}_{ij} - \mathbf{X}\mathbf{s}_{ij}\right\|_2 < \varepsilon, \tag{5}$$

where $\varepsilon > 0$ is the error tolerance, $\mathbf{s}_{ij} \in \mathrm{R}^N$ is the representation coefficient vector of $\mathbf{x}_{ij}$, whose element associated with $\mathbf{x}_{ij}$ is equal to zero. The L_1-minimization problem (5) can be efficiently solved by SLEP [16].

In real-world applications, images are usually corrupted or occluded. Thus, a training sample $\mathbf{x}_{ij}$ can be rewritten as $\mathbf{x}_{ij} = \mathbf{X}\mathbf{s}_{ij} + \mathbf{e}$, in which $\mathbf{e}$ is the corresponding error. The problem (5) can be reformulated as follows:

$$\min_{\boldsymbol{\omega}_{ij}\geq 0} \quad \left\|\boldsymbol{\omega}_{ij}\right\|_1 \quad \text{s.t.} \quad \left\|\mathbf{x}_{ij} - [\mathbf{X}, \mathbf{I}]\,\boldsymbol{\omega}_{ij}\right\|_2 < \varepsilon, \tag{6}$$

where $\boldsymbol{\omega}_{ij} = \left[\mathbf{s}_{ij}^{\mathrm{T}}, \mathbf{e}_{ij}^{\mathrm{T}}\right]^{\mathrm{T}} \in \mathrm{R}^{N+D}$, and $\mathbf{I} \in \mathrm{R}^{D\times D}$ is the identity matrix. Having obtained the optimal representation vector $\tilde{\boldsymbol{\omega}}_{ij} = \left[\tilde{\mathbf{s}}_{ij}^{\mathrm{T}}, \tilde{\mathbf{e}}_{ij}^{\mathrm{T}}\right]^{\mathrm{T}}$