Neurocomputing 155 (2015) 32–42

Contents lists available at ScienceDirect

Neurocomputing

journal homepage: www.elsevier.com/locate/neucom

Kernel Low-Rank Representation for face recognition

Hoangvu Nguyen [a,b,d], Wankou Yang [a,b,c,*], Fumin Shen [e], Changyin Sun [a,b]

[a] School of Automation, Southeast University, Nanjing 210096, China
[b] Key Lab of Measurement and Control of Complex Systems of Engineering, Ministry of Education, Southeast University, Nanjing 210096, China
[c] Jiangsu Key Laboratory of Image and Video Understanding for Social Safety, Nanjing University of Science and Technology, Nanjing, 210094, China
[d] Faculty of Industrial Engineering, Tien Giang University, My Tho 860000, Viet Nam
[e] School of Computer Science and Engineering, University of Electronic Science and Technology of China, Chengdu 610054, China

ARTICLE INFO

Article history:
Received 19 June 2014
Received in revised form
7 December 2014
Accepted 18 December 2014
Communicated by: Su-Jing Wang
Available online 30 December 2014

Keywords:
Low Rank Representation
Kernel method
Sparse representation
Feature extraction
Classification

ABSTRACT

Face recognition is one of the fundamental problems of computer vision and pattern recognition. Based on the recent success of Low-Rank Representation (LRR), we propose a novel image classification method for robust face recognition, named Low-Rank Representation-based Classification (LRRC). Based on seeking the lowest-rank representation of a set of test samples with respect to a set of training samples, the algorithm has the natural discrimination to perform classification. We also propose a Kernel Low-Rank Representation-based Classification (KLRRC), which is a nonlinear extension of LRRC. KLRRC is firstly utilized to face recognition. By using the kernel tricks, we implicitly map the input data into the kernel feature space associated with a kernel function. We construct a transformation matrix to reduce the dimensionality of the kernel feature space, where LRRC is performed. Experimental results on several face data sets demonstrate the effectiveness and robustness of our methods.

1. Introduction

Face recognition is a widespread research area in computer vision with many useful applications in real life, such as face attendance, access control, security surveillance, etc. Face recognition is also a challenging problem, which suffers from, e.g., aging, occlusion, pose, illumination, and expression variations. Many studies to address this issue have created a huge development in face recognition techniques during the past two decades. The Nearest Neighbor (NN) [1] and the Nearest Subspace (NS) [2] classifier are very simple and are widely used. However, in NN, the test image is classified based on the calculation of its nearest neighbor in the training set. Therefore, it is easily affected by noisy data. NS classifies the test image based on the distances to its nearest subspace, which minimizes the reconstruction error. In recent years, the classification methods based on NN has attracted considerable research interests and several improved classifiers have been developed [3–7]. The Local Mean based Nearest Neighbor (LM-NN) classifier uses the mean of the k Nearest Neighbors within a class as prototype of the class [8]. Zhang and Zhang [9] present an ensemble classifier by combining Rotation Forest and Adaboost. Naseem et al. gave a face recognition method as Linear Regression based Classification (LRC) [10]. Zhou et al. [11] developed a semi-supervised Proximal Support Vector Machine (PSVM) classifier.

In the past a few years, sparse representation has been widely applied in computer vision, including image super-resolution [12], motion segmentation [13], denoising [14], sparse reconstruction [15,16] and learning an l^1-regularized Gaussian Bayesian Network (GBN) [17]. Sparse Representation based Classification (SRC) [18] was proposed as a novel image classification approach and achieved high classification accuracy when the number of features is sufficient [19]. In SRC, it considers each test image as a sparse linear combination of the training samples by solving an l^1-minimization problem. Several important aspects in the application of sparse representation in computer vision and pattern recognition are reviewed in [20]. The SRC algorithm has shown its effectiveness and has been studied extensively in the community. Many researchers have applied the SRC in conjunction with other methods to solve their problems, such as using the Gabor feature for SRC to reduce the computational cost and deal with face occlusion [21], using sparse coding and linear pyramid matching for image classification [22], combining sparse coding with the locality constraints for image classification [23,24], using the selection strategy for test sample based sparse representation [25] and Independent Component Analysis (ICA) [26]. One problem with these methods is that they are still not robust enough under occlusion or disguise in face images.

Recently, Liu et al. established Low-Rank Representation (LRR) [27,28] as an efficient method to perform noise correction and subspace segmentation simultaneously. By solving the nuclear norm minimization problem, in general, LRR aims to seek the

* Corresponding author.
E-mail address: wankou.yang@yahoo.com (W. Yang).

http://dx.doi.org/10.1016/j.neucom.2014.12.051
0925-2312/

lowest-rank representation among all the candidates that represent all vectors as the linear combination of the bases in a dictionary. Based on the recent Low Rank Representation theory, we propose a novel classification scheme named Low-Rank Representation-based Classification (LRRC). We find that, the lowest rank representation of the test samples subject to the training samples brings some discrimination information for classification. Instead of solving the sparse representation by the l^1-minimization problem in SRC, we represent all the test samples with respect to all the training samples by seeking the lowest-rank matrix representation. We show that the Low Rank Representations are as effective as the sparse representations, and LRRC obtain competitive performance with SRC, CRC [29] under many circumstances. In addition, the LRRC method is also fast and robust for face recognition under occlusion and corruption. However, as a linear method in nature, LRRC may still fail to deliver good performance when samples (e.g. face images) are subject to complex nonlinear changes due to large pose, expression or illumination variations.

Kernel methods [30–33] have been widely used to overcome the limitations of linear models for feature extraction and classification because Kernel methods can discover the nonlinear structure of the images. The kernel trick is originally applied to construct nonlinear Support Vector Machines (SVM) [34–36]. An appropriate Mercer kernel [30,31] maps the data in the input space into a high dimensional kernel feature space by a nonlinear mapping. And then a linear classifier or a linear regression method is usually applied in the kernel feature space. Any linear model involving the inner product can be turned into the corresponding nonlinear model by applying the kernel tricks [32,33]. Kernel Principal Component Analysis (KPCA) [37] and Kernel Fisher Discriminant Analysis (KFDA) [38] are developed by applying the kernel method to Principal Component Analysis (PCA) and Fisher Discriminant Analysis (FDA), respectively. Generalized Discriminant Analysis (GDA) is proposed by using the kernel function operator to deal with nonlinear discriminant [39]. Yu et al. presented the kernel Nearest Neighbor (KERNEL-NN) classifier [40]. Yang et al. proposed a kernel Fisher discriminant framework for feature extraction and recognition (CFKD) [41]. Recently, Gao et al. proposed the Kernel Sparse Representation (KSR) [42], Cai et al. proposed the Speed up Kernel Discriminant Analysis (SRKDA) [43], Xu et al. presented a face recognition method that represents and classifies face images in the feature space [44], Zhang et al. [45] proposed the Kernel Sparse Representation based Classification (KSRC), Kang et al. [46] propose a novel kernel SRC framework, and apply it together with local image descriptors to face recognition.

In this paper, we also propose a Kernel Low Rank Representation-based Classification (KLRRC) by applying the kernel trick, which can deal with the nonlinear problems by solving LRRC in the kernel feature space. KLRRC is, by the first time, utilized to face recognition. In KLRRC, the data in the input sample space are implicitly mapped into a high dimensional kernel feature space. The mapping aims to convert the nonlinear relations into linear ones associated with a kernel function. However, since the dimensionality of the kernel feature space is very large and unknown, we cannot directly perform LRRC in this feature space. To solve this problem, the kernel based dimensionality reduction method is required to reduce the dimension of the feature space. By applying the projection matrices in KPCA, we construct a projection matrix to reduce the dimensionality of the kernel feature space into the reduced subspace for KLRRC. The lowest rank representation can be obtained by solving a nuclear norm minimization problem in the reduced subspace.

The rest of this paper is organized as follows. In Section 2, we briefly introduce LRR and the kernel trick. Section 3 presents LRRC, KLRRC and describes the dimensionality reduction method in kernel feature space. Experimental results and comparison with other methods are given in Section 4. Finally, Section 5 concludes this paper.

2. Related works

2.1. Low Rank Representation

Liu et al. [27,28] have developed LRR as a method for recovering subspace structure. It can also be used for unsupervised classification to segment data into linear subspaces. Given a proper dictionary, LRR aims at learning the lowest-rank representation among all the candidates that is capable of representing observed data vectors with linear combination of the dictionary atoms. The regularized rank minimization problem can be formulated as follows:

$$\min_{Z} rank(Z) \quad \text{subject to } X = AZ, \tag{1}$$

where X is an observed data matrix, each column of X is an observation, A is a dictionary, and Z is a lowest-rank representation of data X with respect to a dictionary A. However, the rank function is not convex and thus cannot be efficiently minimized. In the rank minimization problems, the rank function can be replaced by the nuclear norm, (1) corresponds to the following convex optimization problem:

$$\min_{Z} ||Z||_{*} \quad \text{subject to } X = AZ \tag{2}$$

In order to segment the data into the irrespective subspaces, Liu et al. [27] use the observed data matrix X itself as the dictionary. So problem (2) is rewritten as follows:

$$\min_{Z} ||Z||_{*} \quad \text{subject to } X = XZ \tag{3}$$

When the data are noisy, the optimization model of LRR is formulated as follows:

$$\min_{Z,E} ||Z||_{*} + \lambda ||E||_{2,1} \quad \text{subject to } X = XZ + E, \tag{4}$$

where XZ is low rank and E is the associated representation errors (or noises) for the data matrix X. The $||E||_{2,1}$ is adopted to characterize the error term for modeling the sample-specific corruptions, $||E||_F^2$ is chosen for the small Gaussian noise and for the random corruptions $||E||_1$ is an appropriate choice. The parameter $\lambda > 0$ is used to balance the effects of the two parts.

Under some conditions, the solutions of these problems coincide and Low-Rank Representation guarantees exist [47,48]. A number of methods have been proposed for solving the problem involving low-rank recovery, such as an iterative thresholding method [49], Accelerated Proximal Gradient (APG) [50] and Augmented Lagrange Multiplier method (ALM) [51,52]. More specifically, ALM is the state-of-the-art method in terms of speed and accuracy, which is applied to deal with the face recognition task in this paper.

2.2. The kernel trick

The kernel trick is a very well-known technique and applied in machine learning. Some typical methods using the kernel trick are SVM [34–36], KPCA [37], and KFDA [38]. By using the kernel trick, we can easily nonlinearly map data from the original input feature space to a high dimensional kernel feature space, and then solve a linear problem in that space. Let ϕ, $\phi : \mathbb{R}^m \to F$ be a non-linear mapping from $\mathbb{R}^m$ into a higher dimensional feature space F so that for any two points x, y, we have a positive semi-definite kernel matrix $K(x,y)$. A Mercer kernel is a function $k(x,y)$ to map the data with a mapping ϕ into a feature space F:$k(x,y) = \langle \phi(x), \phi(y) \rangle$. It can be shown that using k instead of dot product in input space corresponds to map the data with the mapping ϕ. Some commonly used kernels include polynomial kernels $k(x,y) = (\langle x,y \rangle + c)^d$ and Gaussian kernels $k(x,y) = \exp(-\gamma ||x-y||^2)$, where γ and d are the parameters. Thus, a linear algorithm can easily generalize into a