Pattern Recognition 42 (2009) 2327 – 2334

Contents lists available at ScienceDirect

Pattern Recognition

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

Feature extraction based on Laplacian bidirectional maximum margin criterion

Wankou Yang[a,*], Jianguo Wang[b], Mingwu Ren[a], Jingyu Yang[a], Lei Zhang[c], Guanghai Liu[d]

[a]School of Computer Science & Technology, Nanjing University of Science and Technology, Nanjing 210094, China
[b]Network & Education Center, Tangshan College, Tangshan 063000, China
[c]Dept. of Computing, The Hong Kong Polytechnic University, Hong Kong
[d]School Computer Science and Information Technology, Guangxi Normal University, Guangxi, China

ARTICLE INFO

Article history:
Received 28 July 2008
Received in revised form 2 March 2009
Accepted 9 March 2009

Keywords:
MMC
Bidirectional
Laplacian
Feature extraction
Face recognition

ABSTRACT

Maximum margin criterion (MMC) based feature extraction is more efficient than linear discriminant analysis (LDA) for calculating the discriminant vectors since it does not need to calculate the inverse within-class scatter matrix. However, MMC ignores the discriminative information within the local structures of samples and the structural information embedding in the images. In this paper, we develop a novel criterion, namely Laplacian bidirectional maximum margin criterion (LBMMC), to address the issue. We formulate the image total Laplacian matrix, image within-class Laplacian matrix and image between-class Laplacian matrix using the sample similar weight that is widely used in machine learning. The proposed LBMMC based feature extraction computes the discriminant vectors by maximizing the difference between image between-class Laplacian matrix and image within-class Laplacian matrix in both row and column directions. Experiments on the FERET and Yale face databases show the effectiveness of the proposed LBMMC based feature extraction method.

1. Introduction

Dimensionality reduction is an important research topic in computer vision and pattern recognition fields. The curse of high dimensionality is usually a major cause of limitations of many practical technologies, while the large quantities of features may even degrade the performances of the classifiers when the size of the training set is small compared with the number of features [1]. In the past several decades, many feature extraction methods have been proposed, and the most well-known ones are principle component analysis (PCA) and linear discriminant analysis (LDA) [2].

Un-supervised learning cannot properly model underlying structure and characteristics of different classes. Discriminant features are often obtained by supervised learning. LDA [2] is the traditional approach to learn discriminant subspace. Unfortunately, it cannot be applied directly to small size sample (SSS) problems [3] because the within-class scatter matrix is singular. As we know, face recognition is a typical small size problem. Many works have been reported to use LDA for face recognition. The most popular method, called Fisherface, was proposed by Swets et al. [4] and Belhumeur et al. [5]. In their methods, PCA is first used to reduce the dimension of the original features space to $N-c$, and the classical FLD is then applied to reduce the dimension to d ($d \leqslant c$). Since the smallest $c-1$ projection components are thrown away in the PCA step, some useful discriminatory information may be lost. On the other hand, the PCA step cannot guarantee the transformed within-class scatter matrix be nonsingular. More discussions about PCA and LDA can be found in [6].

To solve the singularity problem, a singular value perturbation can be added to the within-class scatter matrix [7]. A more systematic method is regularized discriminant analysis (RDA) [8]. In RDA, one tries to obtain more reliable estimates of the eigenvalues by correcting the eigenvalue distortion with a ridge-type regularization. Penalized discriminant analysis (PDA) is another regularized version of LDA [9,10]. The goals of PDA are not only to overcome the SSS problem but also to smooth the coefficients of discriminant vectors for better interpretation. The main problem of RDA and PDA is that they do not scale well. In applications such as face recognition, the dimensionality is often more than 10,000. It is not practical for RDA and PDA to process such large covariance matrices.

A well-known null subspace method is the LDA + PCA method [11]. When within-class scatter matrix is of full rank, LDA+PCA only calculates the maximum eigenvectors of $(S_w)^{-1}S_b$ to form the transformation matrix. Otherwise, a two-stage procedure is employed. First, the data are transformed into the null space V_0 of S_w. Second, it maximizes the between-class scatter in V_0. LDA + PCA could be sub-optimal because it maximizes the between-class scatter in the

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

0031-3203/$ - see front matter © 2009 Elsevier Ltd. All rights reserved.
doi:10.1016/j.patcog.2009.03.017

null space of S_w instead of the original input space. Direct LDA is another null space method that discards the null space of S_b [12]. It is achieved by diagonalizing first S_b and then S_w, which is in the reverse order of conventional simultaneous diagonalization procedure. If S_t, instead of S_w, is used in direct LDA, it is actually equivalent to the PCA+LDA. Gao et al. [13] proposed a singular value decomposition (SVD) based LDA approach to solving the single training sample per person problem for face recognition. Zhuang and Dai [14,15] developed an inverse Fisher discriminant analysis (IFDA) method. They modified the procedure of PCA and derived the regular and irregular information from the within-class scatter matrix by a new criterion called inverse Fisher discriminant criterion. Jin et al. [16] proposed the uncorrelated optimal discrimination vectors (UODV) approach which maximizes Fisher criterion simultaneously. To avoid the singularity problem of LDA, Li et al. [17] used the difference of both between-class scatter and within-class scatter as discriminant criterion, called maximum margin criterion (MMC). Since the inverse matrix does not to be constructed, the SSS problem in traditional LDA is alleviated. MMC has the advantages of effectiveness and simplicity.

The above-mentioned methods need to transform the 2D images into 1D vectors. This often leads to the so-called "curse of dimensionality" problem, which is always encountered in SSS cases such as face recognition. The matrix-to-vector transform may also cause the loss of some useful structural information embedding in the original images. To overcome the problems, Yang et al. proposed the 2-dimensional principal component analysis (2DPCA) [18]. 2DPCA is based on 2D image matrices rather than 1D vectors. That is, the image matrix does not need to be transformed into a vector. Instead, the image covariance matrix can be constructed directly from the image matrices, and its eigenvectors are derived for image feature extraction. In contrast to PCA, the size of covariance matrix using 2DPCA is much smaller. As a result, 2DPCA computes the corresponding eigenvectors more quickly than PCA. Inspired by the successful application of 2DPCA to face recognition, 2DLDA was proposed [19–22]. Recently, Zheng et al. investigated the relations between vector-based LDA and matrix-based discriminant analysis [23]. They pointed out that from the bias estimation point of view, 2DLDA might be more stable than 1DLDA.

A drawback of 2DPCA is that it needs more coefficients than PCA for image representation. Thus, 2DPCA needs more memory to store features and costs more time to classify. Zuo et al. proposed bidirectional PCA (BDPCA) [24,25] to solve this problem. BDPCA assumes that the transform kernel of PCA is separable and it is a natural extension of the classical PCA and a generalization of 2DPCA. Inspired by 2DPCA, Gao et al. [26] proposed a sequential row–column independent component analysis (RC-ICA) for face recognition.

Recently, a method based on the local geometrical structure called tensor subspace analysis (TSA) [27] was proposed, which captures an optimal linear approximation to the face manifold in the sense of local isometry. However, the computational convergence of the iterative TSA algorithms is not guaranteed. To address the problem, Tao et al. proposed a tensor discriminant analysis method for feature extraction [28,29]. They proposed a convergent solution to discriminative tensor subspace selection and proved the convergence of it. In [30], Zhang et al. presented a directional multilinear ICA method by encoding the image or high-dimensional data array as a general tensor.

Recent studies have shown that the face images possibly reside on a nonlinear submanifold [31–39]. Many manifold-based learning algorithms have been proposed for discovering the intrinsic low-dimensional embedding of the original data. Among the various methods, the most well-known ones are isometric feature mapping (ISOMAP) [31], local linear embedding (LLE) [32] and Laplacian eigenmap [33]. Experiments have shown that these methods can find perceptually meaningful embedding for facial or digit images and other artificial and real-world data sets. However, how to evaluate the maps they generated on novel test data points remains unclear. He et al. [34,35] proposed the locality preserving projections (LPP), which is a linear subspace learning method derived from Laplacian eigenmap. In contrast to most manifold learning algorithms, LPP possesses a remarkable advantage that it can generate an explicit map. This map is linear and can be easily computed, like PCA and LDA. The objective function of LPP is to minimize the local scatter of the projected data.

Yang et al. [36] developed an unsupervised discriminant projection (UDP) technique for dimensionality reduction. UDP characterizes the local scatter as well as the nonlocal scatter, seeking for a projection that simultaneously maximizes the nonlocal scatter and minimizing the local scatter. Both LPP and UDP do not use the class label information and they are unsupervised methods in nature. Yan et al. proposed the marginal Fisher analysis (MFA) [37,38] and Chen et al. proposed the local discriminant embedding (LDE) [39] for feature extraction and recognition. The two methods are very similar in formulation. Both of them combine locality and class label information to represent the intraclass compactness and interclass separability. MFA and LDE take advantage of the partial structural information of classes and neighborhoods of samples; however, it is difficult to decide the number of nearest neighbors of each sample and the number of shortest pairs from different classes in MFA and LDE. In addition, the region covariance matrix (RCM) lies on the connected Riemannian manifold, instead of the subspace. RCM has many merits and is a natural feature for pattern recognition tasks. Pang et al. kernelized the RCM, formalized the similarity metric using four block matrices and obtained good results on face recognition [40].

In this paper, we propose a Laplacian bidirectional maximum margin criterion (LBMMC) for feature extraction and recognition. We formulate the Laplacian between-class scatter matrix and Laplacian within-class scatter matrix on local patches of the data by the weighted summation of distances based on image matrices. The weighted summation of distances has been successfully used in manifold learning [35,36] and can capture the underlying clustering of samples. The objective function of our proposed method is the trace difference criterion which can be directly solved by generalized eigenvalue decomposition. There is no convergence problem in our proposed method. Wang et al. [41] pointed out that the family objective functions for dimensionality reduction with trace ratio criterion can be generally transformed into the corresponding ratio trace criterion for obtaining a closed-form but approximate solution. They proved the convergence of the projection and gave the global optimality of the trace ratio value. They further extended the method into tensor space [42], but it needs much more computation and may be locally optimal in the tensor space.

Recently, Fu et al. have done some very good work in subspace learning [43–45]. In [43,44], they proposed a new criterion based on the concept of k-nearest-neighbor simplex (kNNS), which is constructed by the k-nearest-neighbors, to determine the class label of a new datum. For feature extraction, they developed a novel subspace learning algorithm, called discriminant simplex analysis (DSA), in which the within-class compactness and between-class separability are both measured by kNNS distance. In another work [45], Fu et al. proposed a new discriminant subspace learning algorithm, called correlation tensor analysis (CTA), incorporating both graph-embedded correlation mapping and discriminant analysis in a Fisher type of learning manner. The correlation metric can estimate the intrinsic angles and distance for the locally isometric embedding, which can deal with the case when Euclidean metric fails. CTA learns multiple interrelated subspaces to obtain a low-dimensional data representation reflecting both class label information and intrinsic geometric structure of the data distribution.