J. Vis. Commun. Image R. 32 (2015) 55–62

Contents lists available at ScienceDirect

J. Vis. Commun. Image R.

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

Two-dimensional principal component analysis based on Schatten p-norm for image feature extraction ☆

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 12 February 2015
Accepted 13 July 2015
Available online 23 July 2015

Keywords:
Schatten p-norm
Frebenius-norm
Two-dimensional principal component analysis (2DPCA)
Dimensionality reduction
Feature extraction
Subspace learning
Image classification
Convex optimization

ABSTRACT

In this paper, we propose a novel Schatten p-norm-based two-dimensional principal component analysis (2DPCA) method, which is named after 2DPCA-Sp, for image feature extraction. Different from the conventional 2DPCA that is based on Frobenius-norm, 2DPCA-Sp learns an optimal projection matrix by maximizing the total scatter criterion based on Schatten p-norm in the low-dimensional feature space. Since p can take different values, 2DPCA-Sp is regarded as a general framework of 2DPCA. We also propose an iterative algorithm to solve the optimization problem of 2DPCA-Sp with $0 < p < 1$, which is simple, effective, and easy to implement. Experimental results on several popular image databases show that 2DPCA-Sp with $0 < p < 1$ is robust to impact factors (e.g. illuminations, view directions, and expressions) of images.

1. Introduction

In many image classification and recognition applications, e.g. face recognition, feature extraction plays a key role due to its contribution to alleviate the so-called "curse of dimensionality" [1] and the computation burden. In the past two decades, researchers have presented a lot of vector-based feature extraction methods, including principal component analysis (PCA) [2], linear discriminant analysis (LDA) [3], locality preserving projection (LPP) [4], margin Fisher analysis (MFA) [5], sparsity preserving projection (SPP) [6], etc. Although vector-based feature extraction methods have been successfully applied in many real-world image classification and recognition applications, they need previously transform image matrices into image vectors. The matrix-to-vector transformations inevitably discard the underlying spatial information of images. This might make vector-based methods not optimal for extracting most representative or discriminative features [7]. Recently, Yu et al. presented some subspace learning and feature extraction methods, such as sparse patch alignment framework (SPAF) [8], adaptive hyper-graph learning [9], multimodal hyper-graph learning [10], high-order distance-based multi-view stochastic learning (HD-MSL) [11], and multi-view subspace learning [12]. These methods have been successfully applied in image clustering, image classification, and web image re-ranking, respectively. Moreover, Liu and Tao [13] introduced multi-view Hessian regularization (mHR) to multi-view semi-supervised learning (mSSL) for image annotation. Xu et al. [14] extended the theory of the information bottleneck (IB) and proposed a large-margin multi-view information bottleneck (LMIB) method, which models the multi-view learning problem by using a communication system with multiple senders, each of which represents one view of the data. Furthermore, they also proposed a multi-view intact space learning algorithm [15] to integrate the encoded complementary information of multiple views of the data.

In order to adequately utilize the underlying spatial information of image, many matrix-based feature extraction methods, such as two-dimensional PCA (2DPCA) [16], two-dimensional LDA (2DLDA) [17], two-directional maximum scatter difference (2DMSD) [18], two-dimensional LPP (2DLPP) [19], and Binary 2DPCA [20], have been developed. Different from vector-based feature extraction methods, matrix-based ones directly treat an image as a 2D matrix rather than as a 1D vector.

It must be pointed out that most of the above mentioned methods are based on L2-norm or Frobenius-norm criterion. The L2- or Frobenius-norm-based methods are prone to influence by outliers because L2- or Frobenius-norm criterion can amplify the effect of outliers. As a result, the presence of outliers may push the projection vectors from the desired directions. It is well known that

☆ This paper has been recommended for acceptance by Yehoshua Zeevi.

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

http://dx.doi.org/10.1016/j.jvcir.2015.07.011
1047-3203/

L1-norm is more robust than L2- or Frobenius-norm [21,22]. Hence, many L1-norm-based methods have been developed for feature extraction. Representative L1-norm-based methods are L1-norm PCA (L1-PCA) [23], R1-PCA [24], PCA-L1 [25], LDA-R1 [26] and LDA-L1 [27,28], etc. Among these L1-norm-based PCA methods [23–25], PCA-L1 is an efficient, robust and rotationally invariant method. As reported by Ref. [25], the optimization technique of PCA-L1 is intuitive, simple, and easy to implement. However, since Ref. [25] solves the L1-norm maximization problem by using a greedy strategy, it is prone to get stuck in local solution. To solve this issue, Nie et al. [29] proposed a robust PCA with non-greedy L1-norm maximization, in which they used an efficient non-greedy algorithm to solve the optimization problem of PCA-L1. Moreover, Kwak [30] proposed several PCA methods based on Lp-norm criterion, which try to seek projections that maximize the general Lp-norm with arbitrary $p > 0$ in the feature space. Following the basic idea of PCA-L1, Li et al. [7] proposed an L1-norm-based 2DPCA (2DPCA-L1) method, which is a robust L1-norm version of conventional 2DPCA. For notational clarity, here we refer to the conventional 2DPCA as 2DPCA-L2. Moreover, Pang et al. [31] presented an L1-norm-based tensor analysis (TPCA-L1) method. As a supervised method, LDA-L1 [27,28] is an effective and robust L1-norm version of conventional LDA [3], which seeks a set of local optimal projection vectors by maximizing the ratio of the L1-norm-based between-class scatter to the L1-norm-based within-class scatter in the feature space.

Over the last few years, Schatten p-norm criterion has attracted much attention in machine learning and pattern recognition fields. Nie et al. [32] proposed a low-rank matrix recovery method based on Schatten p-norm minimization to recover a low-rank matrix with a fraction of its entries arbitrarily corrupted, and derived an efficient algorithm to solve the Schatten$p$-norm-based optimization problem. Furthermore, they also proposed a robust matrix completion method based on joint Schatten p-norm and Lp-norm minimization [33,34], which can better approximate the rank minimization problem and is more robust to the outliers. Luo et al. [35] developed a Schatten p-norm-based matrix regression model for image classification, in which they presented a general framework for solving Schatten p-norm with Lq regularization minimization problem. Gu et al. [36] proposed a discriminative metric based on Schatten p-norm. By analyzing the statistical properties of Schatten p-norm metric, they clearly gave the reason why the differences facial images caused by impact factors (e.g. illuminations, view directions, and expressions) are larger than the differences due to identity variations under Frobenius-norm metric, and declared that Schatten p-norm metric is more robust to impact factors of images when $0 < p \leqslant 1$ [36]. Based on these observations, they proposed a Schatten 1-norm PCA (SPCA) [36] method. Zhang et al. [37] further pointed out that SPCA is just an approximative algorithm, because it imposes a very strict constraint, i.e. the projection matrix must be an orthogonal matrix, on the procedure of maximizing the Schatten 1-norm-based PCA criterion, while the projection matrix is usually a column-rank-deficient matrix in many real-world applications. To solve this issue, they proposed an exact Nuclear-norm-based 2DPCA (N-2DPCA) [37] for image feature extraction. However, both SPCA and N-2DPCA are only concerned with a special case of Schatten p-norm with $p = 1$.

In this paper, we propose a Schatten p-norm-based 2DPCA (2DPCA-Sp) method by maximizing the total scatter criterion based on Schatten p-norm in the low-dimensional feature space. Since choosing different p values can suit for different applications, the proposed 2DPCA-Sp can be regarded as a general framework of 2DPCA. It is easy to know that the conventional 2DPCA-L2 [8] and SPCA [36] (or N-2DPCA [37]) are special cases of 2DPCA-Sp with $p = 2$ and $p = 1$, respectively. Although 2DPCA-Sp is theoretically defined for $0 < p < +\infty$, we define it for $0 < p < 1$ to make it more robust to outliers and insensitive to impact factors of images. To solve the objective of 2DPCA-Sp with $0 < p < 1$, we also derive an efficient iterative algorithm.

The remainder of this paper is organized as follows. Section 2 briefly reviews the conventional 2DPCA. The proposed 2DPCA-Sp is presented in Section 3. Section 4 gives the experimental results. Finally, Section 5 concludes this paper.

2. Conventional 2DPCA

As a classical matrix-based feature extraction method, 2DPCA [16] has been widely used in machine learning and pattern recognition fields. Suppose there are N training image matrices, the ith training image matrix is denoted as $\mathbf{X}_i$, the size of which is $h \times w$. 2DPCA aims at finding a projection matrix $\mathbf{U} = [\mathbf{u}_1, \mathbf{u}_2, \ldots, \mathbf{u}_d] \in \mathbb{R}^{w \times d}$ to transform $\mathbf{X}_i$ into the feature space, i.e.

$$\mathbf{Y}_i = \mathbf{X}_i\mathbf{U}, \tag{1}$$

and meanwhile maximizing the total scatter criterion $f(\mathbf{U})$ in the feature space,

$$f(\mathbf{U}) = \sum_{i=1}^{N}\left\|\mathbf{Y}_i - \overline{\mathbf{Y}}\right\|_F^2, \quad \text{s.t. } \mathbf{U}^T\mathbf{U} = \mathbf{I}_d, \tag{2}$$

where $\overline{\mathbf{Y}} = \sum_{i=1}^{N}\mathbf{Y}_i$ is the mean feature matrix, and $\|\mathbf{A}\|_F$ is the Frobenius-norm of matrix $\mathbf{A}$. By some simple algebra, the total scatter criterion can be rewritten as

$$f(\mathbf{U}) = \mathrm{Tr}\left(\mathbf{U}^T\mathbf{S}_t\mathbf{U}\right), \quad \text{s.t. } \mathbf{U}^T\mathbf{U} = \mathbf{I}_d, \tag{3}$$

where $\mathbf{S}_t$ is the image total scatter matrix

$$\mathbf{S}_t = \sum_{i=1}^{N}(\mathbf{X}_i - \overline{\mathbf{X}})^T(\mathbf{X}_i - \overline{\mathbf{X}}), \tag{4}$$

and $\overline{\mathbf{X}} = \sum_{i=1}^{N}\mathbf{X}_i$ is the mean image matrix.

The purpose of 2DPCA is to seek the optimal projection matrix maximizing the total scatter criterion in the feature space, i.e.

$$\max_{\mathbf{U}} \mathrm{Tr}\left(\mathbf{U}^T\mathbf{S}_t\mathbf{U}\right), \quad \text{s.t } \mathbf{U}^T\mathbf{U} = \mathbf{I}_d. \tag{5}$$

The solution of optimization problem (1) can be obtained by computing the orthogonal eigenvectors of $\mathbf{S}_t$ that correspond to the d largest eigenvalues.

3. Schatten *p*-norm-based 2DPCA

The conventional 2DPCA-L2 tries to find the optimal projection matrix to maximize the total scatter criterion based on Frebenius-norm in the feature space. However, it is well known that the Frebenius-norm metric is sensitive to outliers, which means that the presence of outliers may make the solution of 2DPCA-L2 deviate from the desired solution. Motived by the existing Schatten p-norm-based models [32–37], we propose a general Schatten p-norm-based 2DPCA (2DPCA-Sp) method by maximizing the Schatten p-norm-based total scatter criterion.

3.1. Problem formulation

The extended Schatten p-norm $(0 < p < \infty)$ [32–34] of a matrix $\mathbf{A} \in \mathbb{R}^{n \times m}$ is defined as

$$\|\mathbf{A}\|_{S_p} = \left(\sum_{i=1}^{\min\{n,m\}}\sigma_i^p\right)^{\frac{1}{p}} = \left(\mathrm{Tr}\left(\left(\mathbf{A}\mathbf{A}^T\right)^{\frac{p}{2}}\right)\right)^{\frac{1}{p}}, \tag{6}$$

where σ_i is the i-th singular value of $\mathbf{A}$. It is easy to know that Nuclear-norm and Frobenius-norm are special cases of Schatten p-norm with $p = 1$ and $p = 2$, respectively.