Journal of Visual Communication and Image Representation 55 (2018) 67–79

Contents lists available at ScienceDirect

Journal of Visual Communication and Image Representation

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

Incremental generalized multiple maximum scatter difference with applications to feature extraction☆

Ning Zheng[a], Xin Guo[b], Yun Tie[b,*], Nan Dong[c], Lin Qi[b], Ling Guan[c]

[a] School of Electrical and Communication Engineering, Zhengzhou University of Aeronautics, Zhengzhou, China
[b] School of Information and Engineering, Zhengzhou University, Zhengzhou, China
[c] Electrical and Computer Engineering, Ryerson University, Toronto, Canada

ARTICLE INFO

Keywords:
Feature extraction
Generalized multiple maximum scatter difference
Incremental GMMSD+

ABSTRACT

In this paper, we propose a new algorithm to implement the generalized multiple maximum scatter difference (GMMSD). Due to enhanced features of this algorithm over the original GMMSD, we named it GMMSD+. By employing a different projection from both the range of the between-class scatter matrix and the null space of the within-class scatter matrix, GMMSD+ can divide the centroid vector of each class into two components: intrinsic common component (ICC) and discriminant difference component (DCC), and then automatically discards ICC which contains little discriminative information, while keeping DCC which contains the true discriminative power. Next, we introduce a practical implementation of GMMSD+, which can accurately and efficiently update the discriminant vectors with new training samples incrementally, eliminating the complete re-computation of the training process. Our experiments demonstrate that incremental version of GMMSD+(IGMMSD+) eliminates the complete re-computation of the training process when new training samples are presented, leading to significantly reduced computational cost.

1. Introduction

Feature extraction is a key research issue in pattern recognition and machine learning. Effective feature extraction methods normally lead to superior classification performance. The frequently used algorithms of feature extraction include principal component analysis (PCA) [1], linear discriminant analysis (LDA) [2] and independent component analysis(ICA) [3] and so on. However, PCA does not take into account the class information of training set, LDA often suffers from the so-called small sample size (S^3) problem or singularity problem and ICA is not good at extracting features. To address these problems, many more effective LDA-based algorithms such as PCA + LDA [4], Null-LDA (NLDA) [5], LDA/GSVD [6], LDA/QR [7], and manifold learning methods [8–10] have been proposed in the past 20 years.

Yang et al. [11] proposed a complete LDA (CLDA) to overcome the S^3 problem of LDA. CLDA uses the dual space of within-class scatter matrix S_w, i.e. the range space and null space of S_w, to derive more powerful discriminant vectors than LDA. Discriminant common vector (DCV) [12] is another algorithm proposed to overcome the S^3 problem. Similar to other null space algorithms, DCV utilizes the null space of within-class scatter matrix to obtain the optimal discriminant vectors to avoid the singularity problem implicitly. However, Liu [13] pointed out that DCV cannot always obtain the optimal classification accuracy or the robustness in all applications.

Using the trace ratio objective, the singularity problem of LDA can be addressed since the trace of within-class scatter matrix is always larger than zero even if within-class scatter matrix is singular [14]. Despite the existence of a closed-form solution to the singularity problem, the obtained solution does not necessarily best optimize the corresponding trace ratio optimization algorithm. To address the problem, the maximum scatter difference (MSD) [15] was proposed. MSD utilizes the generalized scatter difference rather than the generalized Raleigh quotient as a class separability measure to avoid the S^3 problem. Multiple maximum scatter difference (MMSD) has been also developed to perform feature extraction from both ranges of between-class scatter matrix and null space of the within-class scatter matrix [16]. Although MMSD improves the speed and performance compared with MSD, it needs to calculate SVD twice and the computational complexity is still high for high dimensional data. To solve this problem, the authors [17] proposed the generalized multiple maximum scatter difference (GMMSD), which employs QR decomposition rather than SVD to obtain the optimal discriminant vectors.

☆ This paper has been recommended for acceptance by Guodong Guo.

* Corresponding author.
E-mail addresses: zhengning34@hotmail.com (N. Zheng), guoxin19880806@163.com (X. Guo), ieytie@zzu.edu.cn (Y. Tie), gdongnan@gmail.com (N. Dong), ielqi@zzu.edu.cn (L. Qi), lguan@ee.ryerson.ca (L. Guan).

https://doi.org/10.1016/j.jvcir.2018.04.012
Received 31 March 2017; Received in revised form 23 October 2017; Accepted 28 April 2018
Available online 22 May 2018
1047-3203/

One common difficulty in using the above mentioned methods is that whenever new training samples are presented, the methods have to completely repeat the training process. When the dimension of training set is very high, both the computation and space complexity grow drastically. Thus, an incremental learning algorithm is highly desired to update the discriminant vectors for the insertion of new training samples. To achieve this, several methods have been proposed to perform incremental PCA (IPCA) [18–20] by updating the eigenspace models. Motivated by the incremental PCA, incremental LDA (ILDA) methods have also been investigated. Ye et al. [7] proposed an incremental version of LDA, namely incremental dimension reduction via QR decomposition (IDR/QR), which allows only one new sample in each time. A further limitation of IDR/QR is that, when several samples are inserted into training set, it cannot perform as well as the batch method. Pang et al. [21] introduced a model for updating the between-class scatter matrix and within-class scatter matrix. Uray et al. proposed an ILDA method that combines discriminative and reconstructive information. Zhao et al. [22] studied an incremental LDA for face recognition where generalized SVD is adopted. To overcome the difficulty of incrementally updating the product of scatter matrices, Yan et al. [23] proposed a modified maximum margin criterion by computing the difference of the between-class and within-class scatter matrices. However, this method may lead to regularization problem of the two scatter matrices. Liu et al. [24] proposed a least square-based incremental method which calculated the least square solution of the batch LDA. However, least square incremental linear discriminant analysis (LS-ILDA) method allows only one new sample to be inserted into training set for each updating instance. Recently, Lu et al. [25] proposed a new implementation of complete LDA, which can make full use of the discriminative information of training set, and extended it to the incremental CLDA method. However, these features are not necessarily the optimal set of discriminant vectors.

In this paper, we propose an improved form of GMMSD, which we call GMMSD+, for feature extraction using the principles of GMMSD. The work begins with the batch version of the proposed method which projects centroid matrix onto the range space of the whitened input data to obtain the discriminant vectors based on the principles of GMMSD. The key feature of GMMSD+ is the division of the centroid vector of each class into two components: intrinsic common component (ICC) and discriminant difference component (DDC). The proposed method is able to automatically discard the ICC which contains little discriminative information, while keeping the DDC which contains the true discriminative power. The discovery lays the foundation for the development of the incremental GMMSD+ by which the discriminant vectors of GMMSD+ are accurately updated in an incremental fashion when new training samples are presented, avoiding the re-computation of the whole GMMSD+. As a result, the computational cost is drastically reduced when new training samples are presented. This feature is critically important in pattern recognition, especially in large scale applications such as analytics of big data.

The rest of this paper is organized as follows. In Section 2, we review briefly related work. The basic GMMSD+ algorithm is proposed and analyzed in Section 3. In Section 4, the incremental version of GMMSD+ is presented, with a thorough discussion and comparison on the computational complexity of GMMSD+ and incremental GMMSD+. Section 5 experimentally demonstrates the performance of the proposed method. We conclude the paper in Section 6.

2. Review of MMSD and GMMSD

Suppose that a dataset given in R^m contains m samples from c classes, and x_i^k is the i-th sample in the k-th class, $k = 1, 2, ..., c$, $i = 1, 2, ..., n_k$, where n_k denotes the sample size of the k-th class, $\sum_{k=1}^{c} n_k = n$. The centroid of the k-th class is defined by $\mu_k = \frac{1}{n_k}\sum_{i=1}^{n_k} x_i^k$ and the total centroid of the dataset is defined by $\mu = \frac{1}{n}\sum_{k=1}^{c}\sum_{i=1}^{n_k} x_i^k$. The between-class scatter matrix S_b and within-class scatter matrix S_w are then defined respectively as follows:

$$S_b = \sum_{k=1}^{c} n_k(\mu_k-\mu)(\mu_k-\mu)^T \tag{1}$$

$$S_w = \sum_{k=1}^{c}\sum_{i=1}^{n_k} (x_i-\mu_k)(x_i-\mu_k)^T \tag{2}$$

Define the matrices

$$H_w = [X_1-\mu_1 e_1^T, X_2-\mu_2 e_2^T, ..., X_c-\mu_c e_c^T] \in \Re^{m\times n} \tag{3}$$

$$H_b = [\sqrt{n_1}(\mu_1-\mu), \sqrt{n_2}(\mu_2-\mu), ..., \sqrt{n_c}(\mu_c-\mu)] \in \Re^{m\times c} \tag{4}$$

$$H_t = [x_1^1-\mu, x_2^1-\mu, ..., x_{n_c}^c-\mu] \in \Re^{m\times n} \tag{5}$$

where $X_k = [x_1^k, x_2^k, ..., x_{n_k}^k]$, $e_k^T = (1, ..., 1) \in \Re^{1\times n_k}$. Then, the three scatter matrices in)(3)–(5) can be expressed as $S_t = S_b + S_w = H_t H_t^T$, $S_w = H_w H_w^T$ and $S_b = H_b H_b^T$.

It is intuitive that if the score of the between-class scatter gets larger and the score of the within-class scatter gets smaller, the projected samples can be separated more easily. Thus, the multi-objective optimization principle is written as [15].

$$\max \frac{w_{v1}^T S_b w_{v1}}{w_{v1}^T w_{v1}} \tag{6}$$

$$\min \frac{w_{v2}^T S_w w_{v2}}{w_{v2}^T w_{v2}} \tag{7}$$

where unit vector w_{v1} and w_{v2} are called weight.

A multi-objective optimization problem is often converted into a single-objective optimization problem by the method of objective combination in which all objectives are combined into one scalar objective. There are two principles for meeting the requirement [16]: multiplicative and additive. The classical LDA method can be obtained by applying multiplicative principles to solve the scalar objective problem. On the other hand, Song et al. [16] applied additive principle to combine the two objectives to obtain the MMSD discriminant criterion which is given below

$$\max_{w_i^T w_j=\delta_{ij}, i,j=1,2,...,l} tr(W^T(S_b-\sigma\cdot S_w)W) \tag{8}$$

where σ is a nonnegative constant which balances the relative merits of the first objective to the second objective, called balance parameter; w_i is the optimal discriminant vector, l is the number of optimal discriminant vectors, and $W = (w_1, w_2, ..., w_l)$.

The criterion in (8) can be written in a more compact form:

$$\max_{w_i^T w_j=\delta_{ij}, i,j=1,2,..,l} \sum_{i=1}^{l} w_i^T(S_b-\sigma\cdot S_w)w_i \tag{9}$$

Obviously, it is expensive, both in computational time and memory requirement, to directly perform SVD on matrix $S_b-\sigma\cdot S_w$ when the dimensionality of the input space is high. To address this problem, Song et al. [16] proposed a simple and effective method to decompose matrix $S_b-\sigma\cdot S_w$, namely MMSD. However, MMSD must perform eigen-decomposition twice to obtain the optimal discriminant vectors. Furthermore, the eigen-decomposition of the total scatter matrix is unmanageable when the dimensionality of input space is sufficiently high. For example, suppose the dimensionality of the input data space is $O(N)$, the complexity is in the range of $O(N^3)$. In the real-world, the dimensionality could easily reach the order of thousands in visual information processing tasks, leading to extremely high computational requirement by MMSD. Therefore, it is highly desirable to improve the efficiency of the method.

Zheng et al. [17] proposed a novel implementation of MMSD to extract the discriminant features, which significantly speeds up the training process without scarifying the quality of performance. The