Pattern Recognition 44 (2011) 1567–1580

Contents lists available at ScienceDirect

Pattern Recognition

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

Kernel discriminant transformation for image set-based face recognition

Wen-Sheng Chu, Ju-Chin Chen, Jenn-Jier James Lien *

Department of Computer Science and Information Engineering, National Cheng Kung University, No. 1, Ta-Hsueh Road, Tainan 701, Taiwan

ARTICLE INFO

Article history:
Received 27 February 2010
Received in revised form
28 January 2011
Accepted 9 February 2011
Available online 16 February 2011

Keywords:
Subspace method
Canonical correlation
Kernel discriminant analysis
Face recognition

ABSTRACT

This study presents a novel kernel discriminant transformation (KDT) algorithm for face recognition based on image sets. As each image set is represented by a kernel subspace, we formulate a KDT matrix that maximizes the similarities of within-kernel subspaces, and simultaneously minimizes those of between-kernel subspaces. Although the KDT matrix cannot be computed explicitly in a high-dimensional feature space, we propose an iterative kernel discriminant transformation algorithm to solve the matrix in an implicit way. Another perspective of similarity measure, namely canonical difference, is also addressed for matching each pair of the kernel subspaces, and employed to simplify the formulation. The proposed face recognition system is demonstrated to outperform existing still-image-based as well as image set-based face recognition methods using the Yale Face database B, Labeled Faces in the Wild and a self-compiled database.

1. Introduction

Automatic face recognition has been an essential requirement in a wide range of computer vision applications, such as human–computer interaction (HCI), content-based image retrieval (CBIR) [1,2], security systems and access control systems [3–5]. However, automatic face recognition is a complex process since the appearance of facial images contains an immense variety of expressions, orientations, lighting conditions, occlusions and so forth. Due to these dramatic variations of facial images, the performance of conventional face recognition systems based on a single testing image [3,6–9] is somewhat limited.

Many researchers have noticed the influence of appearance variations for face recognition, and pursued many studies to tackle this problem. Multiple classifiers [10] and active appearance models (AAM) [11] address facial pose issues. A survey also can be also found in [12]. Illumination compensation [13], quotient image creation [14], and synthesized illuminated exemplars [15] are proposed to tackle the illumination problems. In addition, multiscale facial structure [16] and local contrast enhancement [17] is also explored to recognize faces under varying illumination.

Notwithstanding the contributions of [10,11,13–15,18], a single testing input provides insufficient information to guarantee a reliable recognition performance. Better performance could be obtained from *sets* of testing images since multiple images provide more appearance variances of the input data. Recently, canonical correlation (or principal angles) as a similarity between two image sets has drawn increasing attention. The idea of canonical correlation is to measure the cosine angles between associated basis pairs of linear subspaces that correspond to image sets. Canonical correlation has been an effective representation in capturing image set information [19–23]. However, for nonlinearly distributed patterns, such as facial images with head motions and lighting variations (see Fig. 1 for example), these methods are somewhat limited due to their assumption of linearity.

Since facial images are nonlinearly distributed, this study develops a novel kernel discriminant transformation (KDT) algorithm for image-set based face recognition. The proposed algorithm is based on a canonical difference measure [24]. In spirit, the proposed KDT algorithm bears some resemblances to the DCC method [22] in that they both solve an optimal transformation based on a discriminative criterion and the concept of canonical correlations. However, our method has several significant differences in extending the DCC method to a nonlinear version, i.e., the entire learning process is performed in a high-dimensional feature space. Each input image set is represented as a kernel subspace, instead of a linear subspace, by considering the nonlinearity of KPCA [25].

Since the input here is different from DCC [21], our contribution can be clearly summarized into four-fold: (1) Due to the nonlinearity of facial images, we propose to learn a KDT matrix such that after the transformation, separation within kernel subspaces of the same class are minimized, and at the same time separation between kernel subspaces of different classes are maximized. (2) Although the KDT matrix cannot be explicitly computed in high-dimensional feature space, we develop the KDT algorithm to derive implicit evaluation for the dot products in the

* Corresponding author.
E-mail address: jjlien@csie.ncku.edu.tw (J.J. Lien).

0031-3203/$ - see front matter © 2011 Elsevier Ltd. All rights reserved.
doi:10.1016/j.patcog.2011.02.011

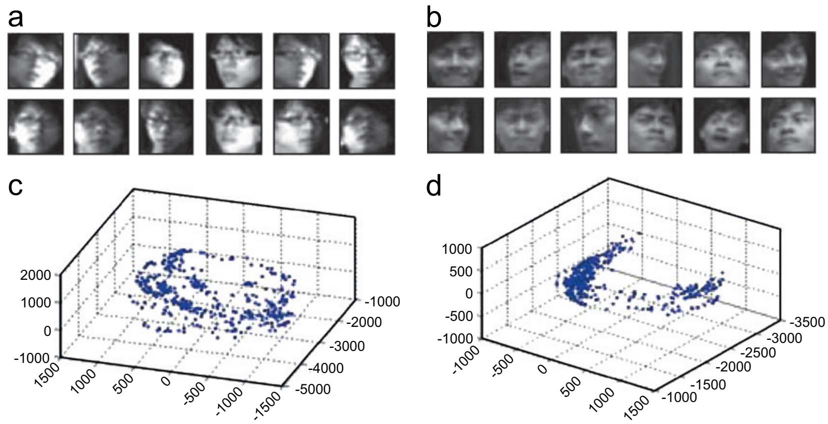

Fig. 1. Typical examples of facial images containing unconstrained head motions with different lighting conditions (a) and facial expressions (b). (c) and (d) show respective nonlinear distribution in the 3D eigenspace.

feature space. (3) We discuss a geometric perspective of canonical correlation, namely *canonical difference*, for measuring the similarity between subspace pairs, and show that the correlation and the difference are in inverse proportion of each other based on simple geometric rules. (4) We provide analysis of computational complexity, bounds, and significance testing for the proposed algorithm.

The rest of this paper is organized as follows: next section reviews previous work related to this paper. An overview of the proposed approach is illustrated in Section 3. Section 4 presents the training procedure, computational analysis and bounds of the proposed KDT algorithm. The testing process is shown in Section 5. We show the experimental results in Section 6, and draw the conclusion in Section 7.

2. Related work

Reviewing the literature, image set-based face recognition approaches mainly fall into two categories: temporal-based and non-temporal approaches. Temporal-based approaches [26–29] recognize human faces by analyzing the connectivity among temporal sequences in video. Non-temporal approaches [20–22,30–36], on the other hand, require no assumption of temporal coherence between facial images, and thus have an advantage that the training database can be arbitrarily expanded instead of being recollected. The proposed method belongs to the later category, therefore we focus on the discussion on non-temporal approaches.

Non-temporal approaches for image set-based face recognition can be further separated into two types [22]: sample-based (nonparametric) and model-based (parametric) methods. In methods of the former type, such as [26], the recognition process involves matching pairwise samples in two image sets, and thus can be time-consuming and sensitive to noise/outliers. By contrast, model-based methods, such as [30,34], assume a preliminary statistical model for each facial image set, and thus requires a strong statistical correlation between the training data and the testing data to ensure satisfactory recognition. To identify that if two facial image sets belong to the same person, the most effective way would be measuring the similarity of the common views of data, which is the idea of canonical correlations.

Yamaguchi et al. propose a set-based face recognition system based on a mutual subspace method (MSM) [19], and yield an improvement compared to image-based approaches that use only one image for testing. Each image set is represented as a linear subspace, and canonical correlations are exploited as the similarity between two image sets. However, MSM does not consider inter-subspace information and therefore limits its discriminative power. Constrained mutual subspace method (CMSM) [20], on the other hand, defines a constrained subspace as differences between all training subspaces, and then measures the canonical correlations in the constrained subspace. Kim et al. [22] report that the performance of CMSM depends on an appropriate dimensionality of the constraint subspace, and propose an alternative scheme to learn a discriminative canonical correlation without specifying dimensionality of target subspace. The recognition performance is shown more robust than both conventional sample-based and model-based methods. Nevertheless, above methods are developed under the linear assumption of input patterns.

Nonlinear extensions of face recognition systems have been widely proposed in [25,37–42]. For these methods, a nonlinear function is applied to map input patterns in the original space to a high-dimensional feature space, where nonlinear input patterns are shown to be more easily classified. The inner products of images in the feature space can be implied by a kernel function in the input space. Regarding the image-based methods, the kernel property is employed in [36] to nonlinearly extend the MSM method [19], denoted as kernel MSM (KMSM). As in the original MSM method, the proposed approach ignores the discriminative information between different classes, and thus restricts its performance. Fukui et al. [31] address the nonlinear version of CMSM in by deriving a representation for the kernel constrained subspace, and justify their performance through an application of 3D object recognition. However, KCMSM inherits the same drawbacks of a appropriate dimensionality for the constrained subspace. In a later study, Fukui and Yamaguchi [32] propose a kernel orthogonal mutual subspace method (KOMSM), where the selection of dimensionality was resolved by orthogonalizing the kernel subspaces before calculating their canonical correlations. Inspired by the above effectiveness of kernel methods, this work proposes a more reliable approach to tackle the problem of image-set based face recognition.

3. Overview of the proposed image set-based face recognition system

Fig. 2 shows the overview of training and testing processes in the proposed image set-based face recognition system. The training process commences by compiling n (typically $n=3$) image sets for each subject. Each image set comprises n_i 20×20-pixel facial images characterized by arbitrary head