Discriminative Learning and Recognition of Image Set Classes Using Canonical Correlations

Tae-Kyun Kim, Josef Kittler, *Member*, *IEEE*, and Roberto Cipolla, *Member*, *IEEE*

Abstract—We address the problem of comparing sets of images for object recognition, where the sets may represent variations in an object's appearance due to changing camera pose and lighting conditions. Canonical Correlations (also known as principal or canonical angles), which can be thought of as the angles between two d-dimensional subspaces, have recently attracted attention for image set matching. Canonical correlations offer many benefits in accuracy, efficiency, and robustness compared to the two main classical methods: parametric distribution-based and nonparametric sample-based matching of sets. Here, this is first demonstrated experimentally for reasonably sized data sets using existing methods exploiting canonical correlations. Motivated by their proven effectiveness, a novel discriminative learning method over sets is proposed for set classification. Specifically, inspired by classical Linear Discriminant Analysis (LDA), we develop a linear discriminant function that maximizes the canonical correlations of within-class sets and minimizes the canonical correlations of between-class sets. Image sets transformed by the discriminant function are then compared by the canonical correlations. Classical orthogonal subspace method (OSM) is also investigated for the similar purpose and compared with the proposed method. The proposed method is evaluated on various object recognition problems using face image sets with arbitrary motion captured under different illuminations and image sets of 500 general objects taken at different views. The method is also applied to object category recognition using ETH-80 database. The proposed method is shown to outperform the state-of-the-art methods in terms of accuracy and efficiency.

Index Terms—Object recognition, face recognition, image sets, canonical correlation, principal angles, canonical correlation analysis, linear discriminant analysis, orthogonal subspace method.

1 Introduction

Many computer vision tasks can be cast as learning problems over vector or image *sets*. In object recognition, for example, a set of vectors may represent a variation in an object's appearance—be it due to camera pose changes, nonrigid deformations, or variation in illumination conditions. The objective of this work is to classify an unknown set of vectors to one of the training classes, each also represented by vector sets. More robust object recognition performance can be achieved by efficiently using set information rather than a single vector or image as input. Examples of pattern sets of an object are shown in Fig. 1.

Whereas most of the previous work on matching image sets for object recognition exploits temporal coherence between consecutive images [20], [21], [11], [22], [28], this study does not make any such assumption. Sets may be derived from sparse and unordered observations acquired by multiple still shots of a three-dimensional object or a long-term monitoring of a scene, as exemplified, e.g., by surveillance systems, where a subject would not face the camera all the time. By this, training sets can be more conveniently augmented in the proposed framework. As this work does not exploit any data semantics explicitly, the proposed method is expected to be applied to many other problems requiring a set comparison.

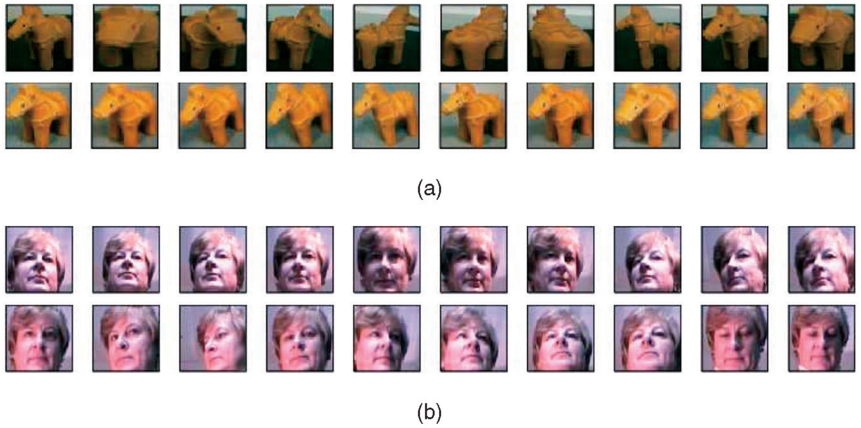
Relevant previous approaches to set matching for set classification can be broadly partitioned into parametric model-based [17], [34] and nonparametric sample-based methods [12], [14]. In the model-based approaches, each set is represented by a parametric distribution function, typically Gaussian. The closeness of the two distributions is then measured by the Kullback-Leibler Divergence (KLD) [6]. Due to the difficulty of parameter estimation under limited training data, these methods easily fail when the training and novel test sets do not have strong statistical relationships.

Rather, more relevant methods for comparing sets are based on matching of pairwise samples of sets, e.g., Nearest Neighbor (NN) and Hausdorff distance matching [12], [14]. The methods are based on the premise that similarity of a pair of sets is reflected by the similarity of the modes (or NN samples) of the two respective sets. This is certainly useful in many computer vision applications where the data acquisition conditions may change dramatically over time. For example, as shown in Fig. 1a, when two sets contain images of an object taken from different views but with a certain overlap in views, global data characteristics of the sets are significantly different making the model-based approaches unsuccessful. To recognize the two sets as the same class, the most effective solution would be to find the common views and measure the similarity of those parts of data. In spite of their rational basis, the nonparametric sample-based methods easily fail as they

- *T.-K. Kim and R. Cipolla are with the Department of Engineering, University of Cambridge, Trumpington Street, Cambridge, CB2 1PZ, UK. E-mail: {tkk22, cipolla}@eng.cam.ac.uk.*
- *J. Kittler is with the Center for Vision, Speech, and Signal Processing, University of Surrey, Guildford, GU2 7XH, UK. E-mail: J.Kittler@surrey.ac.uk.*

Manuscript received 25 July 2006; revised 24 Oct. 2006; accepted 25 Oct. 2006; published online 18 Jan. 2007.
Recommended for acceptance by B. Triggs.
For information on obtaining reprints of this article, please send e-mail to: tpami@computer.org, and reference IEEECS Log Number TPAMI-0554-0706.
Digital Object Identifier no. 10.1109/TPAMI.2007.1037.

0162-8828/07/$25.00 © 2007 IEEE Published by the IEEE Computer Society



(a)

(b)

Fig. 1. Examples of image sets. The sets contain different pattern variations caused by different views and lighting. (a) Two sets (top and bottom) contain images of a 3D object taken from different views but with a certain overlap in their views. (b) Two face image sets (top and bottom) collected from videos taken under different illumination settings. Face patterns of the two sets vary in both lighting and pose.

do not take into account the effect of outliers as well as the natural variability of the sensory data due to the 3D nature of the observed objects. Note also that such methods are very time consuming as they require a comparison of every pair of samples drawn from the two sets.

The above discussion is concerned purely with how to quantify the degree of match between two sets, that is, how to define the similarity of two sets. However, the other important problem in set classification is how to learn discriminative function from training data associated with a given similarity function. To our knowledge, the topic of discriminative learning over sets has not been given proper attention in the literature. In this paper, we interpret the classical Linear Discriminant Analysis (LDA) [14], [7] and its nonparametric variants, Nonparametric Discriminant Analysis (NDA) [19], as techniques of discriminative learning over sets (see Section 2.1). LDA has been recognized as a powerful method for face recognition based on a single face image as input. The methods based on LDA have been widely advocated in the literature [7], [9], [29], [30], [35], [18]. However, note that these methods do not consider multiple input images. When they are directly applied to set classification based on sample matching, they inherit the drawbacks of the classical nonparametric sample-based methods as discussed above.

Most recently, the concept of canonical correlations has attracted increasing attention for image set matching in [36], [8], [23], [24], [37], [27], [39], following the early works [1], [3], [5], [2]. Each set is represented by a linear subspace and the angles between two high-dimensional subspaces are exploited as a similarity measure of two sets (see Section 2.2 for more details). As a method for comparing sets, the benefits of canonical correlations over both parametric distribution-based and sample-based matching, have been noted in our earlier work [36] as well as in [34]. They include efficiency, accuracy, and robustness. This will be discussed and demonstrated in a more detailed and rigorous manner in Section 2.2 and Section 5. A nonlinear extension of canonical correlation has been proposed in [36], [23], [26] and a feature selection scheme for the method in [36]. The Constrained Mutual Subspace Method (CMSM) [24], [37] is closely related to the approach of this paper. In CMSM, a constrained subspace is defined as the subspace in which the entire class population exhibits small variance. The authors showed that the sets of different classes in the constrained subspace had small canonical correlations. However, the principle of CMSM is rather heuristic, especially the process of selecting the dimensionality of the constrained subspace. If the dimensionality is too low, the subspace will be a null space. In the opposite case, the subspace simply captures all the energy of the original data and, thus, cannot play the role of a discriminant function.

This paper presents a novel method of object recognition using image sets, which is based on canonical correlations. The previous conference version [38] has been extended by a more detailed discussion of the key ingredients of the method and the convergence properties of the proposed learning, as well as by reporting the results of additional experiments on face recognition and general object category recognition using the ETH80 [25] database. The main contributions of this paper are as follows: First of all, as a method of comparing sets of images, the benefits of canonical correlations of linear subspaces are explained and evaluated. Extensive experiments comparing canonical correlations with both classical methods (parametric model-based and nonparametric sample-based matching) are carried out to demonstrate these advantages empirically. A novel method of discriminant analysis of canonical correlations is then proposed. A linear discriminant function that maximizes the canonical correlations of within-class sets and minimizes the canonical correlations of between-class sets is defined, by analogy to the optimization concept of LDA. The linear mapping is found by a novel iterative optimization algorithm. Image sets transformed by the discriminant function are then compared by canonical correlations. The discriminative capability of the proposed method is shown to be significantly better than both, the method [8] that simply aggregates canonical correlations and the kNN method applied to image vectors transformed by LDA. Interestingly, the proposed method exhibits very good accuracy as well as other attractive properties: low computational matching cost and simplicity of feature selection. The proposed iterative solution is further compared with classical orthogonal subspace method (OSM) [4], devised to make different subspaces orthogonal to each other. As canonical correlations are only determined up to rotations within subspaces, the canonical correlations of subspaces of between-class sets can be minimized by orthogonalizing those subspaces. To our knowledge, the close relationship of the orthogonal subspace method and canonical correlations has not been noted before. It is also interesting to see that OSM has a close affinity to CMSM. The proposed method and OSM are assessed experimentally on diverse object recognition problems: faces with arbitrary motion under different lighting, general 3D objects observed from different view points, and the ETH80 general object category database. The new techniques are shown to outperform the state-of-the-art methods, including OSM/CMSM and a commercial face recognition software, in terms of accuracy and efficiency.

The paper is organized as follows: The relevant background methods are briefly reviewed and discussed in Section 2. Section 3 highlights the problem of discriminant analysis over sets and presents a novel iterative solution. In Section 4, the orthogonal subspace method is explained and related to both the proposed method and the prior art. The experimental results and their discussion are presented in Section 5. Conclusions are drawn in Section 6.

2 Key Ingredients of the Proposed Learning

2.1 Parametric/Nonparametric Linear Discriminant Analysis

Assume that a data matrix $\mathbf{X}=\{\mathbf{x}_1,\mathbf{x}_2,\ldots,\mathbf{x}_M\}\in\mathbb{R}^{N\times M}$ is given, where $\mathbf{x}_i\in\mathbb{R}^N$ is a N-dimensional column vector obtained by raster-scanning an image. Each vector belongs to one of object classes denoted by C_i. Classical linear discriminant analysis (LDA) finds a transformation $\mathbf{T}\in\mathbb{R}^{N\times n}(n\leq N)$ which maps a vector $\mathbf{x}$ to $\widetilde{\mathbf{x}}=\mathbf{T}^T\mathbf{x}\in\mathbb{R}^n$ such that the transformed data have maximum separation between classes and minimum separation within classes. The between-class and within-class scatter matrices in LDA [7] are given by $\mathbf{B}=\sum_c M_c(\mathbf{m}_c-\mathbf{m})(\mathbf{m}_c-\mathbf{m})^T$, $\mathbf{W}=\sum_c\sum_{\mathbf{x}\in C_c}(\mathbf{x}-\mathbf{m}_c)(\mathbf{x}-\mathbf{m}_c)^T$, where $\mathbf{m}_c$ denotes the class mean, $\mathbf{m}$ is the global mean of the entire sample set and M_c denotes the number of samples in class c. With the assumption that all classes have Gaussian distributions with equal covariance matrix, $trace(\mathbf{B})$ and $trace(\mathbf{W})$ measure the scatter of vectors in the between-class and within-class populations, respectively. A nonparametric form of these scatter matrices is also proposed in [19] with the definition of the between-class and within-class neighbors of a sample $\mathbf{x}_i\in C_c$ given by

$$\mathbf{B}=\frac{1}{M}\sum_{i=1}^{M}w_i(\Delta_i^B)(\Delta_i^B)^T,\quad \mathbf{W}=\frac{1}{M}\sum_{i=1}^{M}(\Delta_i^W)(\Delta_i^W)^T,\tag{1}$$

where $\Delta_i^B=\mathbf{x}_i-\mathbf{x}_i^B$, $\Delta_i^W=\mathbf{x}_i-\mathbf{x}_i^W$,

$$\mathbf{x}^B=\{\mathbf{x}'\in\overline{C}_c\mid\|\mathbf{x}'-\mathbf{x}\|\leq\|\mathbf{z}-\mathbf{x}\|,\forall\mathbf{z}\in\overline{C}_c\},$$

and

$$\mathbf{x}^W=\{\mathbf{x}'\in C_c\mid\|\mathbf{x}'-\mathbf{x}\|\leq\|\mathbf{z}-\mathbf{x}\|,\forall\mathbf{z}\in C_c\}.$$

w_i is a sample weight in order to deemphasize samples away from class boundaries. LDA or Nonparametric Discriminant Analysis (NDA) finds the optimal $\mathbf{T}$ which maximizes $trace(\tilde{\mathbf{B}})$ and minimizes $trace(\tilde{\mathbf{W}})$, where $\tilde{\mathbf{B}},\tilde{\mathbf{W}}$ are the scatter matrices of the transformed data. As these are explicitly represented with $\mathbf{T}$ by $\tilde{\mathbf{B}}=\mathbf{T}^T\mathbf{B}\mathbf{T}$, $\tilde{\mathbf{W}}=\mathbf{T}^T\mathbf{W}\mathbf{T}$, the solution $\mathbf{T}$ can be easily obtained by solving the generalized eigen-problem, $\mathbf{B}\mathbf{T}=\mathbf{W}\mathbf{T}\boldsymbol{\Lambda}$, where $\boldsymbol{\Lambda}$ is the eigenvalue matrix.

When we regard the training data of each class as a set, LDA or NDA can be viewed as the discriminant analysis of the vector sets based on similarity of parametric model-based and nonparametric sample-based matching of sets, respectively. In LDA, each set (i.e., a class) is assumed to be normally distributed with equal covariance matrix and these parametric distributions are optimally separated. On the other hand, in NDA, set similarity is measured by the aggregated distance of a certain number of neighboring samples and the separation of the sets is optimized based on this set similarity.

It is also worth noting that the between-class and within-class scatter measures based on pairwise vector-distance in LDA/NDA can be related to pairwise vector-correlation in many pattern recognition problems. The magnitude of a data vector is often normalized so that $|\mathbf{x}|=1$. As $trace(AB)=trace(BA)$ for any matrix A,B and $|\mathbf{x}|=1$, $trace(\mathbf{W})$ in (1) equals $\frac{1}{M}trace(\sum_i 2(1-\mathbf{x}_i^T\mathbf{x}_i^W))$. The problem of minimizing $trace(\mathbf{W})$ can be changed into the maximization of $trace(\mathbf{W}')$ and similarly the maximization of $trace(\mathbf{B})$ into the minimization of $trace(\mathbf{B}')$, where

$$\mathbf{B}'=\sum_i\mathbf{x}_i^T\mathbf{x}_i^B,\quad\mathbf{W}'=\sum_i\mathbf{x}_i^T\mathbf{x}_i^W\tag{2}$$

and $\mathbf{x}_i^B,\mathbf{x}_i^W$ indicate the closest between-class and within-class vectors of a given vector $\mathbf{x}_i$. Note the weight w_i is omitted for simplicity and the total number of training sets M does not change the direction of the desired components. We now see the optimization problem of classical NDA defined by correlations of pairwise vectors. Rather than dealing with *correlations* of every pair of *vectors*, in the proposed method, we exploit *canonical correlations* of pairwise *linear subspaces* of sets (see Section 3.1 for the proposed problem formulation). By resorting to canonical correlations, the proposed method overcomes the shortcomings of both classical model-based and sample-based approaches in set comparison.

2.2 Definition and Solution of Canonical Correlations

Canonical correlations, which are cosines of principal angles $0\leq\theta_1\leq\ldots\leq\theta_d\leq(\pi/2)$ between any two d-dimensional linear subspaces $\mathcal{L}_1$ and $\mathcal{L}_2$ are uniquely defined as:

$$\cos\theta_i=\max_{\mathbf{u}_i\in\mathcal{L}_1}\max_{\mathbf{v}_i\in\mathcal{L}_2}\mathbf{u}_i^T\mathbf{v}_i\tag{3}$$

subject to $\mathbf{u}_i^T\mathbf{u}_i=\mathbf{v}_i^T\mathbf{v}_i=1,\mathbf{u}_i^T\mathbf{u}_j=\mathbf{v}_i^T\mathbf{v}_j=0,i\neq j$. There are various ways to solve this problem. They are all equivalent but the Singular Value Decomposition (SVD) solution [2] is more numerically stable than the others, as the number of free parameters to estimate is smaller. A comparison with the method called MSM [8] is given in Appendix A. The SVD solution is as follows: Assume that $\mathbf{P}_1\in\mathbb{R}^{N\times d}$ and $\mathbf{P}_2\in\mathbb{R}^{N\times d}$ form unitary orthogonal bases for two linear subspaces, $\mathcal{L}_1$ and $\mathcal{L}_2$. Let the SVD of $\mathbf{P}_1^T\mathbf{P}_2\in\mathbb{R}^{d\times d}$ be

$$\mathbf{P}_1^T\mathbf{P}_2=\mathbf{Q}_{12}\boldsymbol{\Lambda}\mathbf{Q}_{21}^T\quad s.t.\quad\boldsymbol{\Lambda}=diag(\sigma_1,\ldots,\sigma_d),\tag{4}$$

where $\mathbf{Q}_{12}^T\mathbf{Q}_{12}=\mathbf{Q}_{21}^T\mathbf{Q}_{21}=\mathbf{Q}_{12}\mathbf{Q}_{12}^T=\mathbf{Q}_{21}\mathbf{Q}_{21}^T=\mathbf{I}_d$. Canonical correlations are the singular values and the associated canonical vectors, whose correlations are defined as canonical correlations, are given by

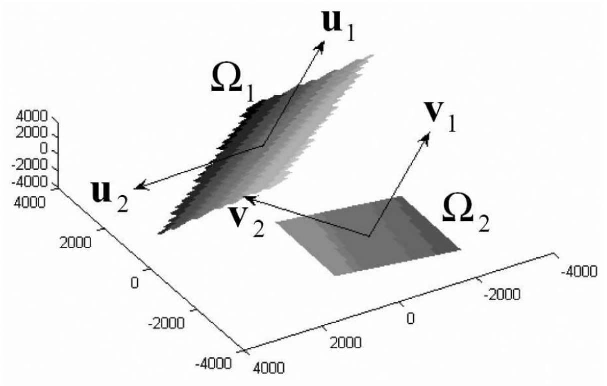


Fig. 2. Conceptual illustration of canonical correlations. Two sets are represented as linear subspaces which are planes here. Canonical vectors on the planes are found to yield maximum correlations. In a two-dimensional subspace case, the second canonical vectors $\mathbf{u}_2, \mathbf{v}_2$ are automatically determined to be perpendicular to the first ones.

$$\mathbf{U} = \mathbf{P}_1\mathbf{Q}_{12} = [\mathbf{u}_1, \ldots, \mathbf{u}_d], \mathbf{V} = \mathbf{P}_2\mathbf{Q}_{21} = [\mathbf{v}_1, \ldots, \mathbf{v}_d]. \quad (5)$$

Canonical vectors are orthonormal in each subspace and $\mathbf{Q}_{12}, \mathbf{Q}_{21}$ can be seen as rotation matrices of $\mathbf{P}_1, \mathbf{P}_2$. The concept is illustrated in Fig. 2.

Intuitively, the first canonical correlation tells us how close are the closest vectors from two subspaces. Similarly, the higher canonical correlations tell us about the proximity of vectors of the two subspaces in other dimensions (perpendicular to the previous ones) of the embedding space. See Fig. 3 for the canonical vectors computed from the sample image sets given in Fig. 1. The common modes (views and/or illuminations) of the two different sets of the same objects are well captured by the first few canonical vectors found. Each canonical vector of one set is very similar to the corresponding canonical vector of the other set despite the data changes across the sets. Also, the canonical vectors of different dimensions represent different variations of the patterns. Compared with the parametric distribution-based matching and the NN matching of samples, this concept is more robust as it effectively places a uniform prior over the subspace of possible pattern variations: Note that a set of high-dimensional vectors of each object is well confined to a characterisitc low-dimensional subspace which retains most of the energy of the set. Then, the proposed subspace-based matching is invariant to the pattern variations subject to the subspaces. On the other hand, the parametric distribution-based matching is highly sensitive to small shift or rotation of the distributions, while the NN matching to small noises or a few outliers contained in the sets. The complexity of the canonical correlation-based method is also very low only requiring SVD of a dxd-dimensional matrix.

3 Discriminant-Analysis of Canonical Correlations (DCC)

As shown in Fig. 3, canonical correlations of two different image sets of the same object acquired in different conditions proved to be a promising measure of similarity of the two sets. This suggests that by matching based on image sets one could achieve a robust solution to the problem of object recognition even when the observation of the object is subject to extensive data variations. However, it is further required to suppress the contribution to similarity of canonical vectors of two image sets due to common environmental conditions (e.g., in lightings, view points, and backgrounds) rather than object identities. The optimal discriminant function is proposed to transform image sets so that canonical correlations of within-class sets are maximized while canonical correlations of between-class sets are minimized in the transformed data space.

3.1 Problem Formulation

Assume m sets of vectors are given as $\{\mathbf{X}_1, \ldots, \mathbf{X}_m\}$, where $\mathbf{X}_i$ describes a data matrix of the ith set containing observation vectors (or images) in its columns. Each set belongs to one of object classes denoted by C_i. A d-dimensional linear subspace of the ith set is represented by an orthonormal basis matrix $\mathbf{P}_i \in \mathbb{R}^{N \times d}$ s.t. $\mathbf{X}_i\mathbf{X}_i^T \simeq \mathbf{P}_i\mathbf{\Lambda}_i\mathbf{P}_i^T$, where $\mathbf{\Lambda}_i, \mathbf{P}_i$ are the eigenvalue and eigenvector matrices of the d largest eigenvalues, respectively, and N denotes the vector dimension. We define a transformation matrix $\mathbf{T} = [\mathbf{t}_1, \ldots, \mathbf{t}_n] \in \mathbb{R}^{N \times n}$, where $n \leq N, |\mathbf{t}_i| = 1$ s.t. $\mathbf{T} : \mathbf{X}_i \rightarrow \mathbf{Y}_i = \mathbf{T}^T\mathbf{X}_i$. The matrix $\mathbf{T}$ transforms images so that the transformed image sets are class-wise more discriminative using canonical correlations.

Representation. Orthonormal basis matrices of the subspaces of the transformed data are obtained from the previous matrix factorization of $\mathbf{X}_i\mathbf{X}_i^T$:

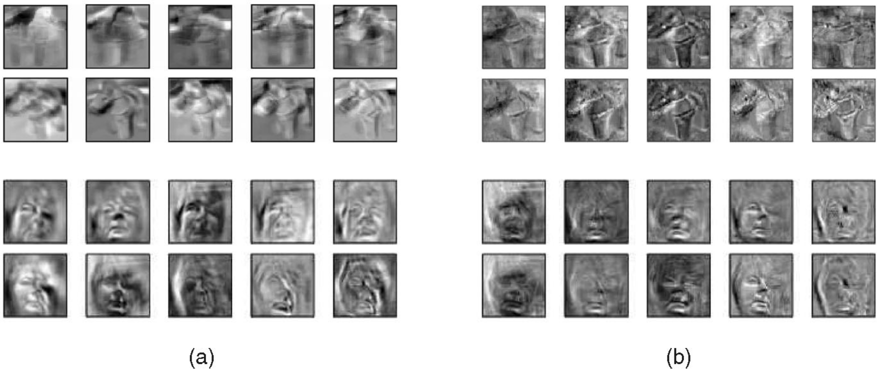


Fig. 3. Principal components versus canonical vectors. (a) The first five principal components computed from the four image sets shown in Fig. 1. The principal components of the different image sets are significantly different. (b) The first five canonical vectors of the four image sets, which are computed for each pair of the two image sets of the same object. Every pair of canonical vectors (each column) $\mathbf{U}, \mathbf{V}$ well captures the common modes (views and illuminations) of the two sets containing the same object. The pairwise canonical vectors are quite similar. The canonical vectors of different dimensions $\mathbf{u}_1, \ldots, \mathbf{u}_5$ and $\mathbf{v}_1, \ldots, \mathbf{v}_5$ represent different pattern variations, e.g., in pose or lighting.

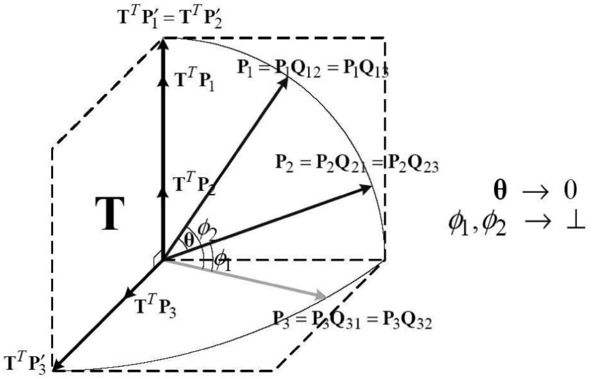

$\mathbf{T}^T\mathbf{P}'_1 = \mathbf{T}^T\mathbf{P}'_2$
$\mathbf{T}^T\mathbf{P}_1$
$\mathbf{P}_1 = \mathbf{P}_1\mathbf{Q}_{12} = \mathbf{P}_1\mathbf{Q}_{13}$
$\mathbf{T}^T\mathbf{P}_2$
$\mathbf{P}_2 = \mathbf{P}_2\mathbf{Q}_{21} = \mathbf{P}_2\mathbf{Q}_{23}$
$\mathbf{T}$
θ
ϕ_2
ϕ_1
$\theta \rightarrow 0$
$\phi_1, \phi_2 \rightarrow \perp$
$\mathbf{T}^T\mathbf{P}_3$
$\mathbf{P}_3 = \mathbf{P}_3\mathbf{Q}_{31} = \mathbf{P}_3\mathbf{Q}_{32}$
$\mathbf{T}^T\mathbf{P}'_3$

Fig. 4. Conceptual illustration of the proposed method. Here, the three sets represented by the basis vector matrices $\mathbf{P}_i, i = 1, \ldots, 3$ are drawn. We assume that the two sets $\mathbf{P}_1, \mathbf{P}_2$ are within-class sets and the third one is coming from the other class. Canonical vectors $\mathbf{P}_i\mathbf{Q}_{ij}, i = 1, \ldots, 3, j \neq i$ are equivalent to basis vectors $\mathbf{P}_i$ in this simple drawing where each set occupies a one-dimensional space. Basis vectors are projected on the discriminative subspace by $\mathbf{T}$ and normalized such that $|\mathbf{T}^T\mathbf{P}'| = 1$. Then, the principal angle of within-class sets, θ becomes zero and the angles of between-class sets, ϕ_1, ϕ_2 are maximized.

$$\mathbf{Y}_i\mathbf{Y}_i^T = (\mathbf{T}^T\mathbf{X}_i)(\mathbf{T}^T\mathbf{X}_i)^T \simeq (\mathbf{T}^T\mathbf{P}_i)\mathbf{\Lambda}_i(\mathbf{T}^T\mathbf{P}_i)^T. \tag{6}$$

Except when $\mathbf{T}$ is an orthogonal matrix, $\mathbf{T}^T\mathbf{P}_i$ is not generally an orthonormal basis matrix. Note that canonical correlations are only defined for orthonormal basis matrices of subspaces (4). Any orthonormal components of $\mathbf{T}^T\mathbf{P}_i$ now defined by $\mathbf{T}^T\mathbf{P}'_i$ can represent an orthonormal basis matrix of the transformed data. See Section 3.2 for details.

Set Similarity. The similarity of any two transformed data sets represented by $\mathbf{T}^T\mathbf{P}'_i$, $\mathbf{T}^T\mathbf{P}'_j$ is defined as the sum of canonical correlations by

$$F_{ij} = \max_{\mathbf{Q}_{ij},\mathbf{Q}_{ji}} \mathrm{tr}(M_{ij}), \tag{7}$$

$$M_{ij} = \mathbf{Q}_{ij}^T\mathbf{P}_i'^T\mathbf{T}\mathbf{T}^T\mathbf{P}'_j\mathbf{Q}_{ji} \text{ or } \mathbf{T}^T\mathbf{P}'_j\mathbf{Q}_{ji}\mathbf{Q}_{ij}^T\mathbf{P}_i'^T\mathbf{T}, \tag{8}$$

as $\mathrm{tr}(AB) = \mathrm{tr}(BA)$ for any matrix A, B. $\mathbf{Q}_{ij}, \mathbf{Q}_{ji}$ are the rotation matrices similarly defined in the SVD solution of canonical correlations (4) with the two transformed subspaces.

Discriminant Function. The discriminative function (or matrix) $\mathbf{T}$ is found to maximize the similarities of any pairs of within-class sets while minimizing the similarities of pairwise sets of different classes. Matrix $\mathbf{T}$ is defined with the objective function J by

$$\mathbf{T} = \arg\max_{\mathbf{T}} J = \arg\max_{\mathbf{T}} \frac{\sum_{i=1}^m \sum_{k\in W_i} F_{ik}}{\sum_{i=1}^m \sum_{l\in B_i} F_{il}}, \tag{9}$$

where the indices are defined as $W_i = \{j|\mathbf{X}_j \in C_i\}$ and $B_i = \{j|\mathbf{X}_j \notin C_i\}$. That is, the two index sets W_i, B_i denote, respectively, the within-class and between-class sets for a given set of class i, by analogy to [19]. See Fig. 4 for the concept of the proposed problem. In the discriminative subspace represented by $\mathbf{T}$, canonical correlations of within-class sets are to be maximized and canonical correlations of between-class sets to be minimized.

3.2 Iterative Learning

The optimization problem of $\mathbf{T}$ involves the variables $\mathbf{Q}, \mathbf{P}'$ as well as $\mathbf{T}$. As the other variables are not explicitly represented by $\mathbf{T}$, a closed form solution for $\mathbf{T}$ is hard to find. We propose an iterative optimization algorithm. Specifically, we compute an optimal solution for one of the three variables at a time by fixing the other two and repeating this for a certain number of iterations. Thus, the proposed iterative optimization is comprised of the three main steps: normalization of $\mathbf{P}$, optimization of matrices $\mathbf{Q}$, and $\mathbf{T}$. Each step is explained below.

Normalization. The matrix $\mathbf{P}_i$ is normalized to $\mathbf{P}'_i$ for a fixed $\mathbf{T}$ so that the columns of $\mathbf{T}^T\mathbf{P}'_i$ are orthonormal. QR-decomposition of $\mathbf{T}^T\mathbf{P}_i$ is performed s.t. $\mathbf{T}^T\mathbf{P}_i = \mathbf{\Phi}_i\mathbf{\Delta}_i$, where $\mathbf{\Phi}_i \in \mathbb{R}^{N\times d}$ is the orthonormal matrix composed by the first d columns and $\mathbf{\Delta}_i \in \mathbb{R}^{d\times d}$ is the $d \times d$ invertible upper-triangular matrix. From (6), $\mathbf{Y}_i = \mathbf{T}^T\mathbf{P}_i\sqrt{\mathbf{\Lambda}}_i = \mathbf{\Phi}_i\mathbf{\Delta}_i\sqrt{\mathbf{\Lambda}}_i$. As $\mathbf{\Delta}_i\sqrt{\mathbf{\Lambda}}_i$ is still an upper-triangular matrix, $\mathbf{\Phi}_i$ can represent an orthonormal basis matrix of the transformed data $\mathbf{Y}_i$. As $\mathbf{\Delta}_i$ is invertible,

$$\mathbf{\Phi}_i = \mathbf{T}^T(\mathbf{P}_i\mathbf{\Delta}_i^{-1}) \quad \rightarrow \quad \mathbf{P}'_i = \mathbf{P}_i\mathbf{\Delta}_i^{-1}. \tag{10}$$

Computation of rotation matrices Q. Rotation matrices $\mathbf{Q}_{ij}$ for every i, j are obtained for a fixed $\mathbf{T}$ and $\mathbf{P}'_i$. The correlation matrix M_{ij} defined in the left of (8) can be conveniently used for the optimization of $\mathbf{Q}_{ij}$, as it has $\mathbf{Q}_{ij}$ outside of the matrix product. Let the SVD of $\mathbf{P}_i'^T\mathbf{T}\mathbf{T}^T\mathbf{P}'_j$ be

$$\mathbf{P}_i'^T\mathbf{T}\mathbf{T}^T\mathbf{P}'_j = \mathbf{Q}_{ij}\mathbf{\Lambda}\mathbf{Q}_{ji}^T, \tag{11}$$

where $\mathbf{\Lambda}$ is a singular matrix and $\mathbf{Q}_{ij}, \mathbf{Q}_{ji}$ are orthogonal rotation matrices. Note that the matrices which are Singular-Value decomposed have only d^2 elements.

Computation of T. The optimal discriminant transformation matrix $\mathbf{T}$ is computed for given $\mathbf{P}'_i$ and $\mathbf{Q}_{ij}$ by using the definition of M_{ij} in the right of (8) and (9). With $\mathbf{T}$ being on the outside of the matrix product M_{ij}, it is convenient to solve for. The discriminative function is found by

$$\mathbf{T} = \max_{arg\mathbf{T}} \mathrm{tr}(\mathbf{T}^T\mathbf{S_b}\mathbf{T})/\mathrm{tr}(\mathbf{T}^T\mathbf{S_w}\mathbf{T}), \tag{12}$$

$$\mathbf{S_b} = \sum_{i=1}^m \sum_{l\in B_i} (\mathbf{P}'_l\mathbf{Q}_{li} - \mathbf{P}'_i\mathbf{Q}_{il})(\mathbf{P}'_l\mathbf{Q}_{li} - \mathbf{P}'_i\mathbf{Q}_{il})^T,$$

$$\mathbf{S_w} = \sum_{i=1}^m \sum_{k\in W_i} (\mathbf{P}'_k\mathbf{Q}_{ki} - \mathbf{P}'_i\mathbf{Q}_{ik})(\mathbf{P}'_k\mathbf{Q}_{ki} - \mathbf{P}'_i\mathbf{Q}_{ik})^T,$$

where $B_i = \{j|\mathbf{X}_j \notin C_i\}$ and $W_i = \{j|\mathbf{X}_j \in C_i\}$. Note that no loss of generality is incurred from (9) as

$$A^TB = \mathbf{I} - 1/2 \cdot (A - B)^T(A - B),$$

where $A = \mathbf{T}^T\mathbf{P}'_i\mathbf{Q}_{ij}, B = \mathbf{T}^T\mathbf{P}'_j\mathbf{Q}_{ji}$. The solution $\{\mathbf{t}_i\}_{i=1}^n$ is obtained by solving the following generalized eigenvalue problem: $\mathbf{S_b}\mathbf{t} = \lambda\mathbf{S_w}\mathbf{t}$. When $\mathbf{S_w}$ is nonsingular, the optimal $\mathbf{T}$ is computed by eigen-decomposition of $(\mathbf{S_w})^{-1}\mathbf{S_b}$. Note also that the proposed learning can avoid a singular case of $\mathbf{S_w}$ by preapplying PCA to data similarly with the Fisherface method [7] and it can be speeded up by using a small number

TABLE 1
Proposed Iterative Algorithm for Finding $\mathbf{T}$, Which Maximizes Class Separation in Terms of Canonical Correlations

```
Algorithm 1. Discriminant-analysis of Canonical Correlations (DCC)
  Input: All $\mathbf{P}_i \in \mathbb{R}^{N\times d}$     Output: $\mathbf{T} \in \mathbb{R}^{N\times n}$
1. $\mathbf{T} \leftarrow \mathbf{I}_N$
2. Do iterate the following:
3.   For all $i$, do QR-decomposition: $\mathbf{T}^T\mathbf{P}_i = \mathbf{\Phi}_i\mathbf{\Delta}_i \rightarrow \mathbf{P'}_i = \mathbf{P}_i\mathbf{\Delta}_i^{-1}$
4.   For every pair $i, j$, do SVD: $\mathbf{P'}_i^T\mathbf{T}\mathbf{T}^T\mathbf{P'}_j = \mathbf{Q}_{ij}\mathbf{\Lambda}\mathbf{Q}_{ji}^T$
5.   Compute $\mathbf{S_b} = \sum_{i=1}^m \sum_{l\in B_i}(\mathbf{P'}_l\mathbf{Q}_{li} - \mathbf{P'}_i\mathbf{Q}_{il})(\mathbf{P'}_l\mathbf{Q}_{li} - \mathbf{P'}_i\mathbf{Q}_{il})^T$,
             $\mathbf{S_w} = \sum_{i=1}^m \sum_{k\in W_i}(\mathbf{P'}_k\mathbf{Q}_{ki} - \mathbf{P'}_i\mathbf{Q}_{ik})(\mathbf{P'}_k\mathbf{Q}_{ki} - \mathbf{P'}_i\mathbf{Q}_{ik})^T$.
6.   Compute eigenvectors $\{\mathbf{t}_i\}_{i=1}^N$ of $(\mathbf{S_w})^{-1}\mathbf{S_b}$,   $\mathbf{T} \leftarrow [\mathbf{t}_1, ..., \mathbf{t}_N]$
7.End
8.$\mathbf{T} \leftarrow [\mathbf{t}_1, ..., \mathbf{t}_n]$
```

of nearest neighboring sets in B_i, W_i similarly with [19]. Canonical correlation analysis for multiple sets [32] is also noteworthy here with regard to fast learning. It may be speeded up by reformulating the between-class and within-class scatter matrices in (12) by the canonical correlation analysis of multiple sets, thus avoiding the computation of the rotation matrices of every pair of image sets in the iterations.

With the identity matrix $\mathbf{I} \in \mathbb{R}^{N\times N}$ as the initial value of $\mathbf{T}$, the algorithm is iterated until it converges to a stable point. A pseudocode for the learning is given in **Algorithm 1** (Table 1). Once $\mathbf{T}$ maximizing the canonical correlations of within-class sets and minimizing those of between-class sets in the training data is found, a comparison of any two novel sets is achieved by transforming them by $\mathbf{T}$ and then computing canonical correlations (see (7)).

3.3 Discussion about Convergence

Although we do not provide a proof of convergence or uniqueness of the proposed optimization process, its convergence to a global maximum was confirmed experimentally. See Fig. 5 for examples of the iterative learning. Each example is for the learning using a different training data set. The value of the objective function J for all cases becomes stable after first few iterations, starting with the initial value $\mathbf{T} = \mathbf{I}$. This fast and stable convergence is very favorable for keeping the learning cost low. Furthermore, as shown at bottom right in Fig. 5, it was observed that the proposed algorithm converged to the same point irrespective of the initial value of $\mathbf{T}$. These results are indicative of the defined criterion being a quadratic convex function with respect to the joint set of variables as well as each individual variable as argued in [15], [10].

For all of the experiments in Section 5, the number of iterations was fixed to five. The proposed learning took about 50 seconds for the face experiments on a Pentium IV PC using nonoptimized Matlab code, while the OSM/CMSM methods took around 5 seconds. Note the learning is performed once in an offliner manner. Online matching by the three recognition methods is highly time-efficient. See the experimental section for more information about the time complexity of the methods.

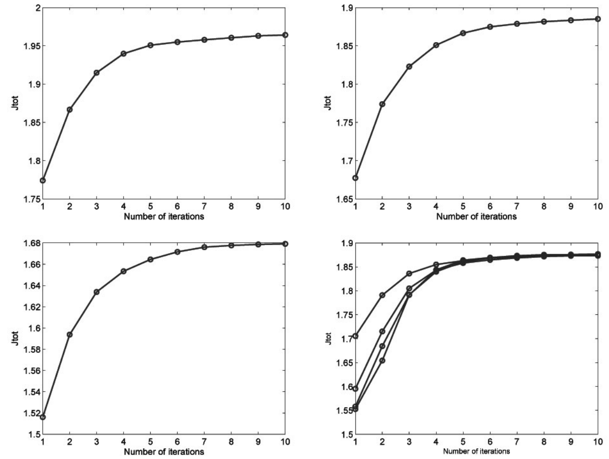


Fig. 5. Convergence characteristics of the optimization: The cost of J of a given training set is shown as a function of the number of iterations. The bottom right shows the convergence to a unique maximum with different random initials of $\mathbf{T}$.

4 Alternative Methods of Discriminative Canonical Correlations for Set Classification

4.1 Orthogonal Subspace Method (OSM)

Orthogonality of two subspaces means that any vector of one subspace is orthogonal to any vector of the other subspace [4]. This requirement is equivalent to that of each basis vector of one subspace being orthogonal to each basis vector of the other. When recalling that canonical correlations are defined as maximal correlations between any two vectors of two subspaces as given in (3), it is very clear that canonical correlations of any two orthogonal subspaces are zeros. Thus, measuring canonical correlations of class specific orthogonal subspaces might be a basis for classifying image sets.

Let us assume that the subspaces of the between-class sets $B_i = \{j|\mathbf{X}_j \notin C_i\}$ of a given data set $\mathbf{X}_i$ are orthogonal to the subspace of the set $\mathbf{X}_i$. If the subspaces are orthogonal, all canonical correlations of those subspaces would also be zero as

$$\mathbf{P}_i{}^T\mathbf{P}_{l\in B_i} = \mathbf{O} \in \mathbb{R}^{d\times d} \quad \rightarrow \quad trace(\mathbf{Q}_{il}^T\mathbf{P}_i{}^T\mathbf{P}_l\mathbf{Q}_{li}) = 0, \tag{13}$$

where $\mathbf{O}$ is a zero matrix and $\mathbf{P}_i$ is a basis matrix of the set $\mathbf{X}_i$. The classical orthogonal subspace method (OSM) [4] has been developed as a method designed to obtain class-specific orthogonal subspaces. The OSM finds the common subspace, which is represented by the basis matrix denoted by $\mathbf{P}_0$, where data sets of different classes become orthogonal. See Appendix B for the details of the OSM solution. By orthogonalizing the subspaces of between-class sets, the discrimination of image sets, in terms of canonical correlations is achieved.

Comparison with the Proposed Solution, DCC. Note that the orthogonality of subspaces is a restrictive condition, at least when the number of classes is large. It is often the case that the subspaces of OSM represented by $\mathbf{P}_i$ and $\mathbf{P}_{l\in B_i}$ are correlated. If $\mathbf{P}_i{}^T\mathbf{P}_l$ has nonzero values, canonical correlations could be much greater than zero as

$$\mathbf{q}_{il}^T\mathbf{P}_i{}^T\mathbf{P}_l\mathbf{q}_{li} \gg 0, \tag{14}$$

Fig. 6. Examples of Cambridge-Toshiba Video Database. The data set contains 100 face classes with varying age, ethnicity, and genders. Each class has about 1,400 images from the 14 image sequences captured under seven different lighting conditions.

where $\mathbf{q}$ is a column of the rotation matrix $\mathbf{Q}$ in the definition of canonical correlations. Generally, the problem of minimizing correlations of basis matrices ${\mathbf{P}_i}^T\mathbf{P}_l$ in OSM is not equivalent to the proposed problem formulation where the canonical correlations $\mathbf{q}_{il}^T{\mathbf{P}_i}^T\mathbf{P}_l\mathbf{q}_{li}$ are minimized. That is, OSM tries orthogonalizatin for all axes of subspaces with equal importance but DCC does this for canonical axes with different importance, revealed by the canonical correlation analysis.

The difference can also be explained in other words with regard to robust input space: Note again that the principal components of $\mathbf{P}$ are sensitive to data changes, whereas canonical vectors $\mathbf{PQ}$ are consistent, as shown in Fig. 3. Thus, the proposed optimization by canonical correlations is expected to be more robust to possible data changes than the OSM solution based on $\mathbf{P}$ (See the discussion in **Further comments on the relationship of DCC, OSM, and CMSM** in Section 5.3.) Moreover, the orthogonal subspace method does not explicitly attempt to maximize canonical correlations of the within-class sets. It combines all examples of a class together. See Appendix B for details.

The similarity matrices of the two methods in Fig. 10 were clearly different reflecting all of the differences noted above. The better accuracy of DCC over OSM was evident when the number of training classes was large or the conditions for obtaining the training and test data were different in the experiments.

4.2 Constrained Mutual Subspace Method (CMSM)

It is worth noting that CMSM [24], [37] can be seen to be closely related to the orthogonal subspace method. For the details of CMSM, refer to Appendix C. CMSM finds the constrained subspace where the total projection operators have small variances. Each class is represented by a subspace which maximally represents the class data variances, then the class subspace is projected into the constrained subspace. The projected data subspace compromises the maximum representation of each class and the minimum representation of a mixture of all the other classes. This is similar in concept with the orthogonal subspace method explained in Appendix B.

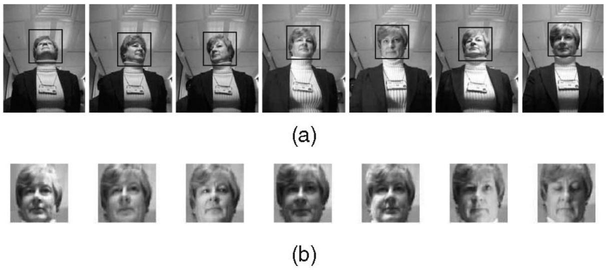


Fig. 7. Example images of the face data sets. (a) Frames of a typical face video sequence with automatic face detection. (b) Face prototypes of seven different illuminations.

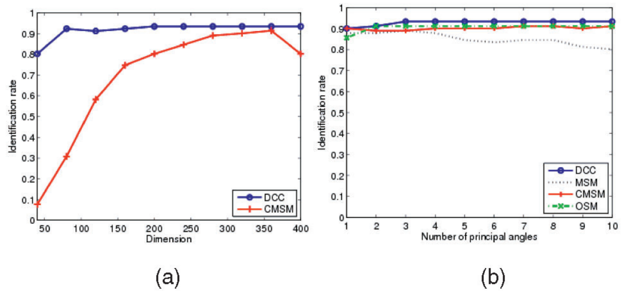
Both methods try to minimize the correlation of between-class subspaces defined by ${\mathbf{P}_i}^T\mathbf{P}_{l\in B_i}$. However, the dimensionality of the constrained subspace of CMSM should be optimized for each application. If the dimensionality is too low, the constrained subspace will be a null space. In the opposite case, the constrained subspace simply retains all the energy of the original data and, thus, cannot play a role as a discriminant function. This dependence of CMSM on the parameter (dimensionality) selection makes it rather empirical. In contrast, there is no need to choose any subspace from the discriminative space represented by the rotation matrix $\mathbf{P}_0$ in the orthogonal subspace method. A full dimension of the matrix can simply be adopted. Note the proposed method, DCC, also exhibited insensitivity to dimensionality, thus being practically, as well as theoretically, very appealing (see Section 5).

5 Experimental Results and Discussion

The proposed method (the code is available at http://mi.eng.cam.ac.uk/~tkk22) is evaluated on various object or object category recognition problems: Using face image sets with arbitrary motion captured under different illuminations, image sets of 500 general objects taken at different views, and the eight general object categories, each of which has several different objects. The task of all of the experiments is to classify an unknown set of vectors to one of the training classes, each also represented by vector sets.

5.1 Database of Face Image Sets

We have collected a database called the *Cambridge-Toshiba Face Video Database* with 100 individuals of varying age and ethnicity and, equally, represented genders, which are shown in Fig. 6. For each person, 14 (seven illuminations × two recordings) video sequences of the person in arbitrary motion were collected. Each sequence was recorded in a different illumination setting for 10 s at 10 fps and at 320×240 pixel resolution. See Fig. 7 for samples from an original image sequence and seven different lightings. Following automatic localization using a cascaded face detector [31] and cropping to a uniform scale of 20×20 pixels, images of faces were histogram equalized. Note that the face localization was performed automatically on the images of uncontrolled quality. Thus, it was not as accurate as any conventional face registration with either manual or automatic eye positions performed on high quality face images. Our experimental conditions are closer to the conditions given for typical surveillance systems.



(a) (b)

Fig. 8. (a) The effect of the dimensionality of the discriminative subspace on the proposed iterative method (DCC) and CMSM. The accuracy of CMSM at 400 is equivalent to that of MSM, a simple aggregation of canonical correlations. (b) The effect of the number of canonical correlations on DCC/MSM/CMSM/OSM.

5.2 Comparative Methods and Parameter Setting

We compared the performance of

- KL-Divergence algorithm (KLD) [17] as a representative parametric model-based method,
- Nonparametric sample-based methods such as k-Nearest Neighbor (kNN) and Hausdorff Distance ($d(S_1, S_2) = \min_{x_1 \in S_1} \max_{x_2 \in S_2} d(x_1, x_2)$) [14] of images transformed by 1) PCA and 2) LDA [7] subspaces, which are estimated from training data similarly to [12],
- Nearest Neighbor (NN) by FaceIt (v.5.0), the commercial face recognition system from Identix, which ranked top overall in the Face Recognition Vendor Test 2000 and 2002 [16], [13],
- Mutual Subspace Method (MSM) [8], which is equivalent to a simple aggregation of canonical correlations,
- Constrained MSM (CMSM) [24], [37] used in a state-of-the-art commercial system called FacePass [40],
- Orthogonal Subspace Method (OSM) [4], and the proposed iterative discriminative learning, DCC.

To compare different algorithms, important parameters of each method were adjusted and the optimal ones in terms of test identification rates were selected. In KLD, 96 percent of data energy was explained by the principal subspace of training data used [17]. In kNN methods, the dimension of PCA subspace was chosen to be 150, which represents more than 98 percent of training data energy (Note that removing the first three components improved the accuracy in the face recognition experiment as similarly observed in [7]). The best dimension of LDA subspace was also found to be around 150. The number of nearest neighbors used was chosen from 1 to 10. In MSM/CMSM/OSM/DCC, the dimension of the linear subspace of each image set represented 98 percent of data energy of the set, which was around 10. PCA was performed to learn the linear subspace of each set in the MSM/CMSM/DCC methods.

Dimension Selection of the Discriminative Subspaces in CMSM/OSM/DCC. As shown in Fig. 8a, CMSM exhibited a high peaking in the the relationship between accuracy and dimensionality of the constrained subspace, whereas the proposed method, DCC, provided constant identification rates regardless of dimensionality of $\mathbf{T}$ beyond a certain point. The best dimension of the constrained subspace of CMSM was found to be at around 360 and was fixed. For DCC, we fixed the dimension at 150 for all experiments (the full dimension can also be conveniently exploited without any feature selection). The full dimension was also used for the rotation matrix $\mathbf{P}_0$ in OSM. Note that the proposed method DCC and OSM do not require any elaborate feature selection and this behavior of DCC/OSM is highly attractive from the practical point of view, compared to CMSM. Without feature selection, the accuracy of CMSM in the full space drops dramatically to the level equivalent to that of MSM, which is a simple aggregation of canonical correlations without any discriminative transformation.

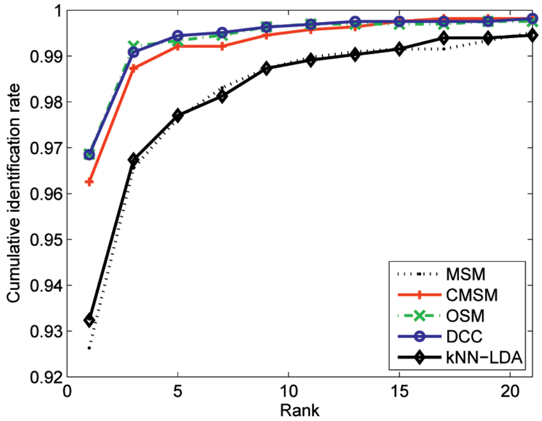


Fig. 9. Cumulative recognition plot for the MSM/kNN-LDA/CMSM/OSM/DCC methods.

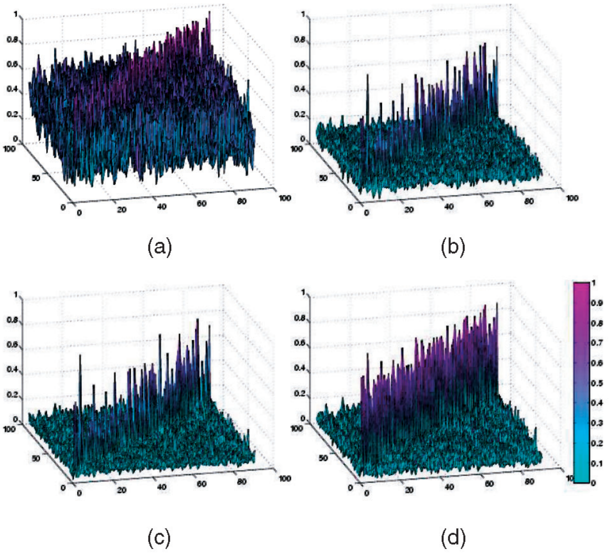


Fig. 10. Similarity matrices for the (a) MSM, (b) CMSM, (c) OSM, and (d) DCC methods. The diagonal and off-diagonal values in the CMSM/OSM/DCC matrix are better distinguished than those of MSM. DCC shows the best separation.

Number of Canonical Correlations. Fig. 8b shows the accuracy of DCC/MSM/CMSM/OSM according to the number of canonical correlations used. Basically, this parameter does not affect the accuracy of the methods as much as the dimension of the discriminative subspace, as shown in Fig. 8a. Overall, the methods, DCC/CMSM/OSM, were shown to be less sensitive to this parameter than MSM, as they exploit their own discriminative transformations. To be more specific, DCC/OSM showed desirably stable curves over this parameter whereas CMSM exhibited

TABLE 2
Evaluation Results

	KLD	HD-PCA	1NN-PCA	10NN-PCA	FaceIt S/W
1st half	0.49±0.14	0.60±0.07	0.95±0.03	0.96±0.03	0.90±0.09
2nd half	0.24±0.13	0.47±0.09	0.71±0.20	0.71±0.21	0.86±0.05
	10NN-LDA	MSM	CMSM	OSM	DCC
1st half	0.98±0.01	0.94±0.03	0.98±0.01	0.98±0.01	0.98±0.01
2nd half	0.87±0.07	0.91±0.02	0.93±0.06	0.94±0.06	0.95±0.04

The mean and standard deviation of recognition rates of different methods. The results are shown separately for the first (easier) and the second parts (more difficult) of the experiments.

more or less fluctuating performance. For simplicity, the number of canonical correlations was fixed to be the same (i.e., this was set as the dimension of linear subspaces of image sets) for all the methods, MSM/CMSM/OSM/DCC.

5.3 Face Recognition Experiments

Training of all the algorithms was performed with data sequences acquired in a single illumination setting and testing with a single other setting. We used 18 randomly selected training/test combinations of the sequences for reporting identification rates. In this experiment, all samples of a training class were drawn from a single video sequence of arbitrary head movement, so they were randomly divided into two sets for the within-class sets in the proposed learning. Note that the proposed method with this random partition still worked well regardless of the number of partitions as exemplified in Table 3. The test recognition rates changed by less than 1 percent for all of the different trials of random partitioning. This may be because no explicit discriminatory information is contained in the randomly partitioned intraclass sets. Rather, DCC is mostly concerned with achieving the maximum possible separation of interclass sets as CMSM/OSM does. In this case, the numerator in the objective function (9) may just help finding the meaningful solution that minimizes the denominator. If samples of a class can be partitioned according to the data semantics, the concept of within-class sets would be more useful and realistic, which is the case in the following experiments. The performance of the evaluated recognition algorithms is shown in Fig. 9 and Table 2. The 18 experiments were divided into two parts according to the degree of difference between the training and the test data of the experiments, which was measured by KL-Divergence between the training and test data. Fig. 9 shows the cumulative recognition rates for the averaged results of all 18 experiments and Table 2 shows the results separately for the first (easier) and the second parts (more difficult) of the experiments.

In Table 2, most of the methods generally had lower recognition rates for the experiments with larger KL-Divergence between the training and test data. The KLD method achieved by far the worst recognition rate. Considering that the illumination conditions varied across data and that the face motion was largely unconstrained, the distribution of within-class face patterns was very broad, making this result unsurprising. In the methods of nonparametric sample-based matching, the Hausdorff-Distance (HD) measure provided far poorer results than the k-Nearest Neighbors (kNN) methods defined in the PCA subspace. 10NN-PCA yielded the best accuracy of the sample-based methods defined in the PCA subspace, which is worse than MSM by 8.6 percent on average. Its performance greatly varied across the experiments. Note that MSM showed robust performance with a large margin over kNN-PCA method under the different experimental conditions. The improvement of MSM over both KLD and HD/kNN-PCA methods was very impressive. The benefits of using canonical correlations over both classical approaches for set classification, which have been explained throughout the previous sections, were confirmed.

TABLE 3
Example Results for Random Partitioning

Number of partitions	2	3	4
exp. 1	93.19±0.46	92.86±0.78	92.75±0.77
exp. 2	95.93±0.53	95.60±0.00	95.71±0.35

The mean and standard deviation (percent) of recognition rates of 10 random trials for two example experiments.

The commercial face recognition software FaceIt (v.5.0) yielded the performance which is in the middle of those of kNN-PCA and kNN-LDA methods on average. Although the NN method using FaceIt is based on individual sample matching, it delivered more robust performance for the data changes (the difference in accuracy between the first half and the second half is not as large as those of kNN-PCA/LDA methods). This is reasonable, considering that FaceIt was trained independently with the training images used for other methods.

Table 2 also gives a comparison for the methods combined with discriminative learning. kNN-LDA yielded a big improvement over kNN-PCA but the accuracy of the method again greatly varied across the experiments. Note that 10NN-LDA outperformed MSM for similar conditions between the training and test sets, but it became noticeably inferior as the conditions changed. It delivered similar accuracy to MSM on average, which is also shown in Fig. 9. The proposed method DCC, CMSM, and OSM constantly provided a significant improvement over both MSM and kNN-LDA methods as shown in Table 2 as well as in Fig. 9.

Further comments on the relationship of DCC, OSM, and CMSM. Note that CMSM/OSM can be considered as measuring correlation between subspaces defined by the basis matrix $\mathbf{P}$ in a simple way which is different from the canonical correlations defined by $\mathbf{PQ}$. In spite of this difference, the accuracy of CMSM/OSM was impressive in this experiment. As explained above, when an ideal solution of CMSM/OSM exists and $\mathbf{Q}$ only provides a rotation within the subspace, the solution of CMSM/OSM can be close to that of the proposed method DCC. However, if class subspaces cannot be made orthogonal to each other, then the direct optimization of canonical correlations offered by DCC is preferred. The novel data space $\mathbf{PQ}$ is

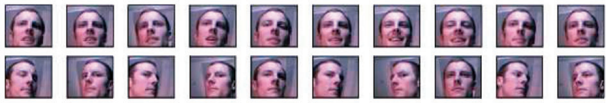


Fig. 11. Example of the two time sets (top and bottom) of a person acquired in a single lighting setting. They contain significant variations in pose and expression.

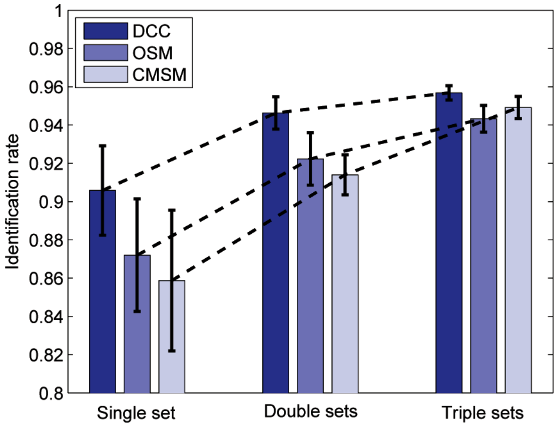
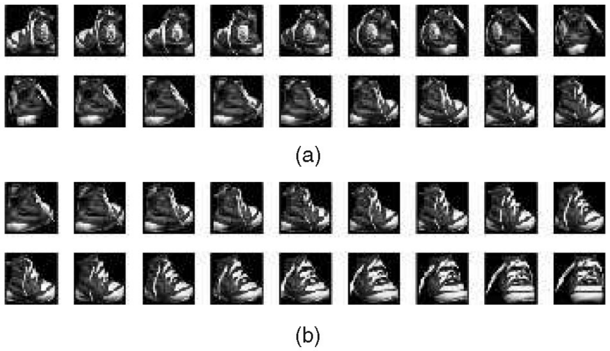


Fig. 12. Recognition rates of the CMSM/OSM/DCC methods when using a single, double, and triple image sets in training.

robust to environmental changes as shown in Fig. 3, making the solution of DCC, which is obtained by directly optimizing $\mathbf{PQ}$ space, also robust. Note that the proposed method was better than CMSM/OSM for the second half of the experiments in Table 2.

The differences of the three methods are clearly apparent from the associated similarity matrices of the training data. We trained the three methods using both training and test sets of the worst experimental case for the methods (see the last two of Fig. 7b), and compared their similarity matrices of the total class data with that of MSM, as shown in Fig. 10. Both OSM and CMSM considerably improved the ability of class discrimination over MSM, but they were still far from optimal compared with DCC for the given data. As discussed above, both of the proposed method, DCC, and OSM are preferable to CMSM as they do not involve the selection of dimensionality of the discriminative subspaces. While the best dimension for CMSM had to be identified with reference to the test results, the full dimension of the discriminative space can simply be adopted for any new test data in the DCC and OSM methods.

We designed another face experiments with more face image sets in the Cambridge-Toshiba face video database. The database involves two sets of videos acquired at different times, each of which consists of seven different illumination sequences for each person. We used one time set for training and the other set for testing thus having more variations between the training and testing (see Fig. 11 for an example of the two sets acquired in the same illumination at different times). Note the training and testing sets in the previous experimental setting were drawn from the same time set. In this experiment using a single illumination set for training, the full 49 combinations of the different lighting settings were exploited. We also increased the number of image sets per each class for training. We randomly drew a combination of different illumination sequences for training and used all seven



(a)

(b)

Fig. 13. ALOI experiment. (a) The training set consists of 18 images taken at 10 degree intervals. (b) Two test sets (each top and bottom row) are shown. Each test set contains nine images at 10 degree intervals, different from the training set.

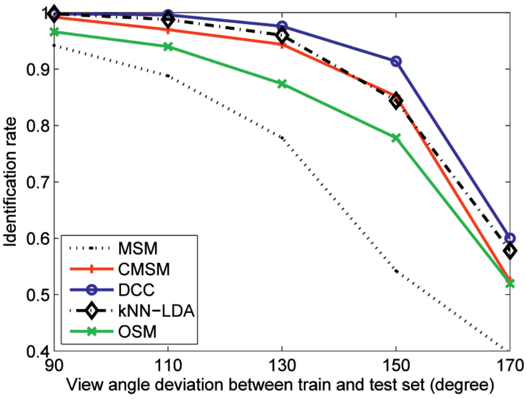


Fig. 14. Identification rates for the five different test sets. The object viewing angles of the test sets differ from those of the training set to a varying extent.

illumination sequences for testing. Ten-fold cross validation was performed for these experiments. Fig. 12 shows the mean and standard deviations of recognition rates of all experiments. The proposed method significantly outperformed OSM/CMSM methods when the test sets were much different from the training sets. These results are consistent with those of the methods in the second part of the experiment in Table 2 (but the difference is much clearer here). Overall, all three methods improved their accuracy by using more image sets in training.

Matching complexity. The complexity of the methods based on canonical correlations (MSM/CMSM/OSM/DCC), $O(d^3)$, is much lower than that of the sample-based matching methods (kNN-PCA/LDA), $O(m^2n)$, where d is the subspace dimension of each set, m is the number of samples of each set and n is the dimensionality of feature vectors, since $d \ll m, n$. In the face experiments, the unit matching time of comparing the two image sets which contain about 100 images is 0.004 for the canonical correlations-based method and 1.1 seconds for the kNN method.

5.4 Experiment on Large-Scale General Object Database

The ALOI database [33] with 500 general object categories taken at different viewing angles provides another experimental data set for the proposed method. Object images were segmented from the simple background and scaled to 20×20 pixel size. A training set and five test sets were set

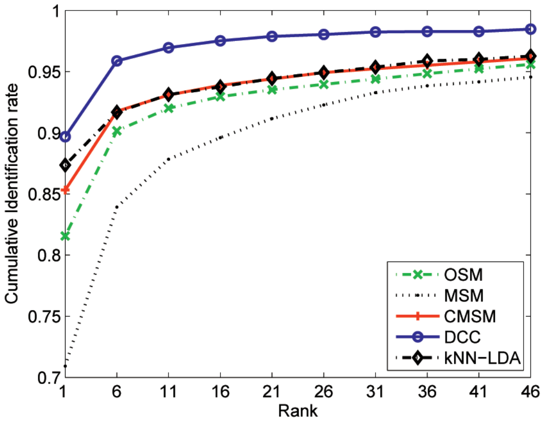


Fig. 15. Cumulative recognition rates of the MSM/kNN-LDA/CMSM/OSM/DCC methods for the ALOI experiment.

up with different viewing angles of the objects, as shown in Figs. 13a and 13b. Note that the pose of all the images in the test sets differed by at least 5 degree from every sample of the training set. The methods of MSM, kNN-LDA, CMSM, and OSM were compared with the proposed method in terms of identification rate. The parameter were selected in the same way as in the face recognition experiment. The dimension of the linear subspace of each image set was fixed to 5, representing more than 98 percent data energy in MSM/CMSM/OSM/DCC methods. The best number of nearest neighbors in the kNN-LDA method was found to be five.

Judging from Fig. 14 and Fig. 15, kNN-LDA yielded better accuracy than MSM in all the cases. This contrasted with the findings in the face recognition experiment. This may have been caused by the somewhat artificial experimental setting. The nearest neighbors of the training and test set differed only slightly due to the five degree pose difference. Please note that the two sets had no changes in lighting and had accurate localization of the objects as well, making the good chance of working for the sample-based method. Further note that the accuracy of MSM could be improved by using only the first canonical correlation, similarly to the results shown in Fig. 8b. Here again, CMSM, OSM, and the proposed method DCC were substantially superior to MSM. Overall, the accuracy of CMSM/OSM was similar to that of kNN-LDA method, as shown in Fig. 15. The proposed iterative method, DCC, constantly outperformed all the other methods including OSM/CMSM as well as kNN-LDA. Please note this experiment involved a larger number of classes, compared with the face experiments. Furthermore, the set of images of the training class had quite different pose distributions from those of the test set. The accuracy of CMSM/OSM methods might be degraded by all these factors, whereas the proposed method is still robust.

5.5 Object Category Recognition Using ETH80 Database

An interesting problem of object category recognition was performed using the public ETH80 database. As shown in Fig. 16, there are eight categories which contain 10 objects each, with 41 images of different views. More details about the database can be found in [25]. We randomly partitioned 10 objects into two sets of five objects for training and testing. In Experiment 1, we used all 41 view images of objects. In Experiment 2, we used all 41 views for training but a random subset of 15 view images for testing. Ten-fold cross-validation was carried out for both experiments. Parameters such as the dimension of the linear subspaces, the number of principal angles and nearest neighbors were selected as in the previous experiment. The dimension of the constrained subspace of CMSM was also best optimized.

From Table 4, it is worth noting that the accuracy of kNN-PCA method is similar (but slightly inferior) to that of the PCA method reported in [25]. Note that we used only five objects per category, in contrast to [25], where nine objects were used for training. The recognition rates for individual object categories also showed similar behavior to those of [25].

As shown in Table 4, the kNN methods were much inferior to the methods based on canonical correlations. The sample-based matching method was very sensitive to the variations in different objects of the same categories, failing in object categorization. The methods using canonical correlations provided much more accurate results. The proposed method (DCC) delivered the best accuracy over

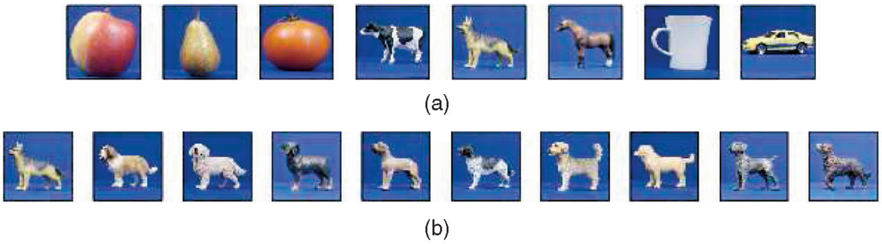


Fig. 16. Object category database (ETH80) contains (a) eight different object categories and (b) 10 different objects for each category.

TABLE 4
Evaluation Results of Object Categorization

	kNN-PCA	kNN-LDA	MSM	CMSM	OSM	DCC
exp.1	0.762±0.21	0.752±0.17	0.865±0.13	0.897±0.10	0.905±0.09	0.917±0.09
exp.2	-	-	-	0.852±0.21	0.865±0.18	0.912±0.13

The mean recognition rate and its standard deviation for all experiments.

all tested methods. The improvement of DCC over CMSM/OSM was bigger in the second experiment where only a subset of images of objects was involved in the testing. Note that this makes the testing set very different from the training set. The major principal components of the image sets are highly sensitive to the variations in pose. The accuracy of CMSM/OSM methods was considerably decreased in the presence of this variation, while the DCC method maintained almost the same accuracy.

6 Conclusions and Future Work

A novel discriminative learning framework has been proposed for set classification based on canonical correlations. It is based on iterative learning which is theoretically and practically appealing. The proposed method has been evaluated on various object and object category recognition problems. The new technique facilitates effective discriminative learning over sets and exhibits an impressive set classification accuracy. It significantly outperformed the KLD method representing a parametric distribution-based matching and kNN methods in both PCA/LDA subspaces as examples of nonparametric sample-based matching. It also largely outperformed the method based on a simple aggregation of canonical correlations.

The proposed DCC method achieved not only better accuracy but also possesses many good properties, compared with CMSM/OSM methods. CMSM had to be optimized a posteriori by feature selection. In contrast, DCC does not need any feature selection. It exhibited a robust performance over a wide range of dimensions of the discriminative subspace as well as the number of canonical correlations used. Although CMSM/OSM delivered a comparable accuracy to DCC in particular cases, in general, it lagged behind the proposed method.

The canonical-correlation-based methods including the proposed method were also shown to be highly time efficient in matching, thus offering an attractive tool for recognition involving a large-scale database.

The interesting research direction of this study is the nonlinear extension of the concept of DCC, which would allow us to capture discriminatory information of image sets contained in higher-order statistics. It may also prove beneficial to make the proposed learning more time-efficient so as to be incrementally updated for new training sets.

Appendix A
Equivalence of SVD Solution to Mutual Subspace Method [8]

In Mutual Subspace Method (MSM), canonical correlations are defined as the eigenvalues of the matrix $\mathbf{P}_1\mathbf{P}_1^T\mathbf{P}_2\mathbf{P}_2^T\mathbf{P}_1\mathbf{P}_1^T \in \mathbb{R}^{N\times N}$, where $\mathbf{P}_i \in \mathbb{R}^{N\times d}$ is a basis matrix of a data set i. The SVD solution in (4) for computing canonical correlations is symmetric. That is,

$$\mathbf{Q}_{12}^T\mathbf{P}_1^T\mathbf{P}_2\mathbf{Q}_{21} = \Lambda,$$

$$\mathbf{Q}_{21}^T\mathbf{P}_2^T\mathbf{P}_1\mathbf{Q}_{12} = \Lambda.$$

By multiplying the above two equations, we obtain

$$(\mathbf{Q}_{12}^T\mathbf{P}_1^T\mathbf{P}_2\mathbf{Q}_{21})(\mathbf{Q}_{21}^T\mathbf{P}_2^T\mathbf{P}_1\mathbf{Q}_{12}) = \Lambda^2$$

$$\rightarrow \mathbf{Q}_{12}^T\mathbf{P}_1^T\mathbf{P}_2\mathbf{P}_2^T\mathbf{P}_1\mathbf{Q}_{12} = \Lambda^2$$

$$\rightarrow \mathbf{P}_1\mathbf{P}_1^T\mathbf{P}_2\mathbf{P}_2^T\mathbf{P}_1\mathbf{P}_1^T = \mathbf{P}_1\mathbf{Q}_{12}\Lambda^2\mathbf{Q}_{12}^T\mathbf{P}_1^T$$

as $\mathbf{Q}_{12}\mathbf{Q}_{12}^T = \mathbf{Q}_{21}\mathbf{Q}_{21}^T = \mathbf{I}$. $\mathbf{P}_1\mathbf{Q}_{12}$ and Λ^2 are the eigenvector matrix and eigenvalue matrix, respectively, of the matrix $\mathbf{P}_1\mathbf{P}_1^T\mathbf{P}_2\mathbf{P}_2^T\mathbf{P}_1\mathbf{P}_1^T$. That is, the canonical correlations of MSM simply assume the square value of the canonical correlations of the SVD solution. Please note that the dimension of the matrix $\mathbf{P}_1^T\mathbf{P}_2 \in \mathbb{R}^{d\times d}$ is relatively low compared with the dimension of $\mathbf{P}_1\mathbf{P}_1^T\mathbf{P}_2\mathbf{P}_2^T\mathbf{P}_1\mathbf{P}_1^T \in \mathbb{R}^{N\times N}$.

Appendix B
OSM Solution

Denote the correlation matrices of the C classes by $\mathbf{C}^1, \dots, \mathbf{C}^C$ and the respective a priori probabilities by $\pi^1, \dots, \pi^C$ [4]. Then, matrix $\mathbf{C}_0 = \sum_{i=1}^{C} \pi^i\mathbf{C}^i$ is the correlation matrix of the mixture of all the classes. Matrix $\mathbf{C}_0$ can be diagonalized by $\mathbf{B}\mathbf{C}_0\mathbf{B}^T = \Lambda$. Denoting $\mathbf{P}_0 = \Lambda^{-1/2}\mathbf{B}$, we have $\mathbf{P}_0\mathbf{C}_0\mathbf{P}_0^T = \mathbf{I}$. Then,

$$\pi^1\mathbf{P}_0\mathbf{C}^1\mathbf{P}_0^T + \dots \pi^C\mathbf{P}_0\mathbf{C}^C\mathbf{P}_0^T = \mathbf{I}.$$

This means that matrices $\pi^i\mathbf{P}_0\mathbf{C}^i\mathbf{P}_0^T$ and $\Sigma_{j\neq i}\pi^j\mathbf{P}_0\mathbf{C}^j\mathbf{P}_0^T$ have the same eigenvectors but the eigenvalues λ_k^i of $\pi^i\mathbf{P}_0\mathbf{C}^i\mathbf{P}_0^T$ and $\overline{\lambda_k^i}$ of $\Sigma_{j\neq i}\pi^j\mathbf{P}_0\mathbf{C}^j\mathbf{P}_0^T$ are related by $\lambda_k^i = 1 - \overline{\lambda_k^i}$. That is, in the space rotated by matrix $\mathbf{P}_0$, the most important basis vectors of class i, which are the eigenvectors of $\pi^i\mathbf{P}_0\mathbf{C}^i\mathbf{P}_0^T$ corresponding to largest eigenvalues, are at the same time the least significant basis vectors for the ensemble of the rest of the classes. Let $\mathbf{P}_i$ be such an eigenvector matrix so that

$$\pi^i{\mathbf{P}_i}^T\mathbf{P}_0\mathbf{C}^i\mathbf{P}_0^T\mathbf{P}_i = \Lambda^i.$$

Then,

$$\Sigma_{j\neq i}\pi^j{\mathbf{P}_i}^T\mathbf{P}_0\mathbf{C}^j\mathbf{P}_0^T\mathbf{P}_i = \mathbf{I} - \Lambda^i.$$

Since every matrix $\pi^j\mathbf{P}_0\mathbf{C}^j\mathbf{P}_0^T$ for all $j \neq i$ is positive semidefinite, $\pi^j{\mathbf{P}_i}^T\mathbf{P}_0\mathbf{C}^j\mathbf{P}_0^T\mathbf{P}_i$ should be a diagonal matrix having smaller elements than $1 - \lambda^i$. If we let $\mathbf{P}_j$ denote the eigenvectors of jth class by $\pi^j\mathbf{P}_0\mathbf{C}^j\mathbf{P}_0^T \approx \mathbf{P}_j\Lambda^j{\mathbf{P}_j}^T$, the matrix ${\mathbf{P}_i}^T\mathbf{P}_j\Lambda^j{\mathbf{P}_j}^T\mathbf{P}_i$ now has small diagonal elements. Accordingly, ${\mathbf{P}_i}^T\mathbf{P}_j$ should have all the elements close to zero. In the ideal case when $\pi^i\mathbf{P}_0\mathbf{C}^i\mathbf{P}_0^T$ has the eigenvalues which are exactly equal to one, the matrix ${\mathbf{P}_i}^T\mathbf{P}_j$ would be a zero matrix for all $j \neq i$. The two subspaces defined by $\mathbf{P}_i, \mathbf{P}_j$ are called orthogonal subspaces. That is, every column of $\mathbf{P}_i$ is perpendicular to every column of $\mathbf{P}_j$.

Note that the OSM method does not exploit the concept of multiple sets in a single class (or within-class sets). The method assumes that all data vectors of a single class i are represented by a single set $\mathbf{P}_i$. From the above, the matrix $\mathbf{P}_0$ could represent an alternative discriminative space where the canonical correlation of between-class sets are minimized. Note that the matrix $\mathbf{P}_0$ is a rotation matrix in its concept and therefore it is a square matrix.

APPENDIX C

CONSTRAINED MUTUAL SUBSPACE METHOD [24]

The constrained subspace $\mathbf{D}$ is spanned by N_d eigenvectors $\mathbf{d}$ of the matrix $\mathbf{G} = \sum_{i=1}^{C} \mathbf{P}_i \mathbf{P}_i^T$ s.t.

$$\mathbf{G}\mathbf{d} = \lambda \mathbf{d},$$

where C is the number of training classes, $\mathbf{P}_i$ is a basis matrix of the original $i\text{th}$ class data, and eigenvector $\mathbf{d}$ corresponds to the N_d smallest eigenvalues. The optimal dimension N_d of the constrained subspace is set experimentally. The subspace $\mathbf{P}_i$ is projected onto $\mathbf{D}$ and the orthogonal components of the projected subspace, normalized to unit length, are obtained as inputs for computing canonical correlations by the method of MSM [8].

ACKNOWLEDGMENTS

The authors gratefully acknowledge the support of Toshiba Corporation, Chevening Scholarship, and Cambridge-Oversea-Trust. They are grateful to the Toshiba Corporation for helping the database construction and motivating this study. Its contributions are also regarded as the first works using canonical correlation analysis for image set-based object recognition problems. Useful discussions and comments from O. Arandjelović and G. Brostow are much appreciated.

REFERENCES

[1] H. Hotelling, "Relations between Two Sets of Variates," *Biometrika,* vol. 28, no. 34, pp. 321-372, 1936.

[2] Å. Björck and G.H. Golub, "Numerical Methods for Computing Angles between Linear Subspaces," *Math. Computation,* vol. 27, no. 123, pp. 579-594, 1973.

[3] T. Kailath, "A View of Three Decades of Linear Filtering Theory," *IEEE Trans. Information Theory,* vol. 20, no. 2, pp. 146-181, 1974.

[4] E. Oja, *Subspace Methods of Pattern Recognition.* Research Studies Press, 1983.

[5] R. Gittins, *Canonical Analysis: A Review with Applications in Ecology.* Springer-Verlag, 1985.

[6] T.M. Cover and J.A. Thomas, *Elements of Information Theory.* Wiley, 1991.

[7] P.N. Belhumeur, J.P. Hespanha, and D.J. Kriegman, "Eigenfaces versus Fisherfaces: Recognition Using Class Specific Linear Projection," *IEEE Trans. Pattern Analysis and Machine Intelligence,* vol. 19, no. 7, pp. 711-720, July 1997.

[8] O. Yamaguchi, K. Fukui, and K. Maeda, "Face Recognition Using Temporal Image Sequence," *Proc. Int'l Conf. Automatic Face and Gesture Recognition,* pp. 318-323, 1998.

[9] W.Y. Zhao, R. Chellappa, and A. Krishnaswamy, "Discriminant Analysis of Principal Components for Face Recognition," *Proc. Int'l Conf. Automatic Face and Gesture Recognition,* pp. 336-341, 1998.

[10] D.D. Lee and H.S. Seung, "Learning the Parts of Objects by Non-Negative Matrix Factorization," *Nature,* vol. 401, no. 6755, pp. 788-791, 1999.

[11] Y. Li, S. Gong, and H. Liddell, "Recognising the Dynamics of Faces Across Multiple Views," *Proc. British Machine Vision Conf.,* pp. 242-251, 2000.

[12] S. Satoh, "Comparative Evaluation of Face Sequence Matching for Content-Based Video Access," *Proc. Int'l Conf. Automatic Face and Gesture Recognition,* pp. 163-168, 2000.

[13] D.M. Blackburn, M. Bone, and P.J. Phillips, *Facial Recognition Vendor Test 2000: Evaluation Report,* 2000.

[14] R.O. Duda, P.E. Hart, and D.G. Stork, *Pattern Classification,* second ed. John Wily and Sons, 2000.

[15] D.D. Lee and H.S. Seung, "Algorithms for Non-Negative Matirx Factorization," *Advances in Neural Information Processing Systems,* pp. 556-562, 2001.

[16] P.J. Phillips, P. Grother, R.J. Micheals, D.M. Blackburn, E. Tabassi, and J.M. Bone, *Facial Recognition Vendor Test 2002: Evaluation Report,* http://www.frvt.org/FRVT2002/, Mar. 2003

[17] G. Shakhnarovich, J.W. Fisher, and T. Darrel, "Face Recognition from Long-Term Observations," *Proc. European Conf. Computer Vision,* pp. 851-868, 2002.

[18] M.-H. Yang, "Kernel Eigenfaces vs. Kernel Fisherfaces: Face Recognition Using Kernel Methods," *Proc. Int'l Conf. Automatic Face and Gesture Recognition,* pp. 215-220, 2002.

[19] M. Bressan and J. Vitria, "Nonparametric Discriminant Analysis and Nearest Neighbor Classification," *Pattern Recognition Letters,* vol. 24, no. 15, pp. 2743-2749, 2003.

[20] K. Lee, M. Yang, and D. Kriegman, "Video-Based Face Recognition Using Probabilistic Apperance Manifolds," *Proc. Computer Vision and Pattern Recognition Conf.,* pp. 313-320, 2003.

[21] S. Zhou, V. Krueger, and R. Chellappa, "Probabilistic Recognition of Human Faces from Video," *Computer Vision and Image Inderstanding,* vol. 91, no. 1, pp. 214-245, 2003.

[22] X. Liu and T. Chen, "Video-Based Face Recognition Using Adaptive Hidden Markov Models," *Proc. Computer Vision and Pattern Recognition Conf.,* pp. 340-345, 2003.

[23] L. Wolf and A. Shashua, "Learning over Sets Using Kernel Principal Angles," *J. Machine Learning Research,* vol. 4, no. 10, pp. 913-931, 2003.

[24] K. Fukui and O. Yamaguchi, "Face Recognition Using Multi-Viewpoint Patterns for Robot Vision," *Proc. Int'l Symp. Robotics Research,* pp. 192-201, 2003.

[25] B. Leibe and B. Schiele, "Analyzing Appearance and Contour Based Methods for Object Categorization," *Proc. Computer Vision and Pattern Recognition Conf.,* pp. 409-415, 2003.

[26] D. Hardoon, S. Szedmak, and J. Shawe-Taylor, "Canonical Correlation Analysis; An Overview with Application to Learning Methods," *Neural Computation,* vol. 16, no. 12, pp. 2639-2664, 2004.

[27] T. Kozakaya, O. Yamaguchi, and K. Fukui, "Development and Evaluation of Face Recognition System Using Constrained Mutual Subspace Method," *IPSJ J.,* vol. 45, no. 3 pp. 951-959, 2004.

[28] A. Hadid and M. Pietikainen, "From Still Image to Video-Based Face Recognition: An Experimental Analysis," *Proc. Sixth IEEE Int'l Conf. Automatic Face and Gesture Recognition,* pp. 813-818, 2004.

[29] M.T. Sadeghi and J.V. Kittler, "Decision Making in the LDA Space: Generalised Gradient Direction Metric," *Proc. Int'l Conf. Automatic Face and Gesture Recognition,* pp. 248-253, 2004.

[30] X. Wang and X. Tang, "Random Sampling LDA for Face Recognition," *Proc. Computer Vision and Pattern Recognition Conf.,* pp. 259-265, 2004.

[31] P. Viola and M. Jones, "Robust Real-Time Face Detection," *Int'l J. Computer Vision,* vol. 57, no. 2, pp. 137-154, 2004.

[32] J. Via, I. Santamaria, and J. Perez, "Canonical Correlation Analysis (CCA) Algorithms for Multiple Data Sets: Application to Blind SIMO Equalization," *Proc. 13th European Signal Processing Conf.,* 2005.

[33] J.M. Geusebroek, G.J. Burghouts, and A.W.M. Smeulders, "The Amsterdam Library of Object Images," *Int'l J. Computer Vision,* vol. 61, no. 1, pp. 103-112, Jan. 2005.

[34] O. Arandjelović, G. Shakhnarovich, J. Fisher, R. Cipolla, and T. Darrell, "Face Recognition with Image Sets Using Manifold Density Divergence," *Proc. Computer Vision and Pattern Recognition Conf.,* pp. 581-588, 2005.

[35] T.-K. Kim and J. Kittler, "Locally Linear Discriminant Analysis for Multimodally Distributed Classes for Face Recognition with a Single Model Image," *IEEE Trans. Pattern Analysis and Machine Intelligence,* vol. 27, no. 3, pp. 318-327, Mar. 2005.

[36] T.-K. Kim, O. Arandjelović, and R. Cipolla, "Learning over Sets Using Boosted Manifold Principal Angles (BoMPA)," *Proc. British Machine Vision Conf.,* pp. 779-788, 2005.

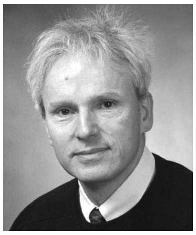
[37] M. Nishiyama, O. Yamaguchi, and K. Fukui, "Face Recognition with the Multiple Constrained Mutual Subspace Method," *Proc. Audio and Video-Based Biometric Person Authentication,* pp. 71-80, 2005.

[38] T.-K. Kim, J. Kittler, and R. Cipolla, "Learning Discriminative Canonical Correlations for Object Recognition with Image Sets," *Proc. European Conf. Computer Vision,* pp. 251-262, 2006.

[39] K. Fukui, B. Stenger, and O. Yamaguchi, "A Framework for 3D Object Recognition Using the Kernel Constrained Mutual Subspace Method," *Proc. Asian Conf. Computer Vision,* pp. 315-324, 2006.

[40] Toshiba Corporation "Facepass," http://www.toshiba.co.jp/rdc/mmlab/tech/w31e.htm, 2005.

Tae-Kyun Kim received the BSc and MSc degrees from the Department of Electrical Engineering and Computer Scince at the Korea Advanced Institute of Science and Technology (KAIST) in 1998 and 2000, respectively. He is a PhD student in the Machine Intelligence Laboratory at the University of Cambridge. He worked as a research staff member at Samsung Advanced Institute of Technology, Korea during 2000-2004. His research interests include computer vision, statistical pattern classification, and machine learning. He regularly reviews for the *IEEE Transactions on Pattern Analysis and Machine Intelligence* (*TPAMI*) and was the Korea delegate of MPEG-7. The joint proposal of face descriptor of Samsung and NEC, for which he developed the main algorithms, was accepted as the international standard of ISO/IEC JTC1/SC29/WG11.



Josef Kittler is a professor of machine intelligence and the director of the Centre for Vision, Speech, and Signal Processing at the University of Surrey. He has worked on various theoretical aspects of pattern recognition and image analysis, and on many applications including personal identity authentication, automatic inspection, target detection, detection of microcalcifications in digital mammograms, video coding and retrieval, remote sensing, robot vision, speech recognition, and document processing. He has coauthored the book *Pattern Recognition: A Statistical Approach* (Prentice-Hall) and published more than 500 papers. He is a member of the editorial boards of *Image and Vision Computing*, *Pattern Recognition Letters*, *Pattern Recognition and Artificial Intelligence*, *Pattern Analysis and Applications*, and *Machine Vision and Applications*. He is a member of the IEEE.

Roberto Cipolla received the BA degree (engineering) from the University of Cambridge in 1984 and the MSE degree (electrical engineering) from the University of Pennsylvania in 1985. In 1991, he was awarded the DPhil degree (computer vision) from the University of Oxford. His research interests are in computer vision and robotics and include the recovery of motion and 3D shape of visible surfaces from image sequences, visual tracking and navigation, robot hand-eye coordination, algebraic and geometric invariants for object recognition and perceptual grouping, novel man-machine interfaces using visual gestures, and visual inspection. He has authored three books, edited six volumes, and coauthored more than 200 papers. He is a member of the IEEE.