Difference Subspace and Its Generalization for Subspace-Based Methods

Kazuhiro Fukui, *Member, IEEE* and Atsuto Maki, *Member, IEEE*

Abstract—Subspace-based methods are known to provide a practical solution for image set-based object recognition. Based on the insight that local shape differences between objects offer a sensitive cue for recognition, this paper addresses the problem of extracting a subspace representing the difference components between class subspaces generated from each set of object images independently of each other. We first introduce the *difference subspace* (DS), a novel geometric concept between two subspaces as an extension of a difference vector between two vectors, and describe its effectiveness in analyzing shape differences. We then generalize it to the *generalized difference subspace* (GDS) for multi-class subspaces, and show the benefit of applying this to subspace and mutual subspace methods, in terms of recognition capability. Furthermore, we extend these methods to kernel DS (KDS) and kernel GDS (KGDS) by a nonlinear kernel mapping to deal with cases involving larger changes in viewing direction. In summary, the contributions of this paper are as follows: 1) a DS/KDS between two class subspaces characterizes shape differences between the two respectively corresponding objects, 2) the projection of an input vector onto a DS/KDS realizes selective visualization of shape differences between objects, and 3) the projection of an input vector or subspace onto a GDS/KGDS is extremely effective at extracting differences between multiple subspaces, and therefore improves object recognition performance. We demonstrate validity through shape analysis on synthetic and real images of 3D objects as well as extensive comparison of performance on classification tests with several related methods; we study the performance in face image classification on the Yale face database B+ and the CMU Multi-PIE database, and hand shape classification of multi-view images.

Index Terms—Subspace method, mutual subspace method, canonical angles, difference subspace, 3D object recognition

---

1 INTRODUCTION

SUBSPACE-BASED methods [1], [2], [3], which provide a practical solution for pattern recognition in image sets, have been a central research topic for object recognition in computer vision. This is because a set of images of an object can be effectively modeled by a low-dimensional subspace of a high-dimensional vector space [4], [5], [6], [7], [8].

The recognition task in this framework is to compare an input image or a set of input images to the image sets of reference class objects, and classify the (set of) input images into the correct class. It is assumed that each object region is roughly segmented and normalized with respect to size in advance of classification [5], [7], [9], [10], [11], [12], [13], [14]. In the recognition process, an image of a 3D object, represented by $w{\times}h$ pixels, is treated as a vector in a $w{\times}h$-dimensional vector space. A set of images of the object is thus represented as a low-dimensional subspace of the vector space. Such a subspace is typically generated by applying to the image set either the Karhunen–Loève expansion (also known as principal component analysis (PCA)), or Gram–Schmidt orthogonalization.

Subspaces considered in this paper implicitly contain information about the shape of a 3D object. When the image set of a 3D object with a convex Lambertian surface under a fixed viewpoint and several different illumination conditions is considered, the subspace generated from the image set is called an *illumination subspace*. It contains explicit information about the shape and albedo of the object in the form of their inner products [4], [5], [6], [7], [8]. If an object has uniform albedo, its illumination subspace contains shape information only. This implies that structural similarity between two illumination subspaces is related to shape similarity between the objects represented by those subspaces [10], [15], [16].

The structural similarity between two subspaces can be mathematically measured by a set of canonical (or principal) angles between them [17], [18]. A canonical angle is a natural extension of the angle between two vectors. All canonical angles between perfectly coincident subspaces are zero; all canonical angles between orthogonal subspaces are 90 degree.

One of the most popular methods of classification by canonical angles is the mutual subspace method (MSM) [9]. MSM is a natural extension of the classic and well-known subspace method (SM) [1], [2], [3], whose input is one vector. In contrast to SM, MSM classifies on the basis of the canonical angles between an input subspace and the object class subspaces. These subspaces are generated from the sets of images of the unknown object and those of the reference class objects, respectively. The characteristics of SM and MSM make them both excellent at handling multiple classes by their nature, unlike binary classifiers including support vector machines. This ability to handle multiple classes is extremely important for multi-class applications such as facial recognition. Both SM and MSM generate a

- *K. Fukui is with the Department of Computer Science, University of Tsukuba, Tsukuba, Japan. E-mail: kfukui@cs.tsukuba.ac.jp.*
- *A. Maki is with the School of Computer Science and Communication, KTH Royal Institute of Technology, Stockholm, Sweden. E-mail: atsuto@kth.se.*

subspace for each class; this also makes them quite different from the Eigenface [19] and Fisherface methods [20].

Despite the strengths of SM and MSM, their ability to directly classify objects is still limited in the sense that the class subspaces are generated independently of each other [3]. Although each subspace is a good representation, in terms of least mean square approximation of the distribution of the training images, there is no reason to assume a priori that these representations are the optimal subspaces for classification. When distinguishing between similar objects, a measure of similarity based on such naively chosen subspaces will be inherently unreliable, because it captures just the global shape similarity between the objects. Our insight is that local shape differences between objects should be a more stable and sensitive measure for dealing with objects having subtle differences.

In this paper, we propose a framework to extract local shape differences between distinct objects and then perform a classification using those differences. To simplify the following discussion, we primarily consider the case that all objects under consideration have uniform albedo. In this case, since the shape information of each object is well modeled by the corresponding illumination subspace, the shape difference between two objects is well captured by the *difference components* between the corresponding illumination subspaces. If desired, this framework can be extended to objects that do not have uniform albedo by simultaneously considering differences in shape and albedo.

The problem we address here is, therefore, how to extract the subspace corresponding to the difference components of a pair of original subspaces. To address the problem, we first geometrically define the concept of *difference subspace* (DS), which represents the difference components between two subspaces; these components are based on the canonical angles. A DS is a natural extension of the difference vector concept to a pair of subspaces. Therefore, the projection of a vector or a subspace onto a DS enables us to extract difference components between two class subspaces from the vector or subspace.

To handle more than two class subspaces, we generalize the concept of a DS to a *generalized difference subspace* (GDS), which contains the difference components among the class subspaces. For this generalization, we first show that a DS between two subspaces can be defined analytically by using the two projection matrices of the subspaces. From this analytical definition, we can then systematically generate a GDS. The projection on a GDS allows us to extract difference components between multiple class subspaces. In the context of 3D object recognition, this projection extracts shape differences between 3D objects.

Interestingly, in addition to the above qualitative understanding, the properties of GDS projection can be understood quantitatively: it expands the canonical angles between different class subspaces towards an orthogonal relationship. In this respect, GDS projection is closely related to the orthogonalization method proposed by Fukunaga and Koontz [21], [22].

Specifically, our approach to enhancing the classification ability of SM and MSM is to incorporate GDS projection into the framework of SM or MSM. See Fig. 1, which shows the concept. We call the enhanced SM and MSM *constrained*
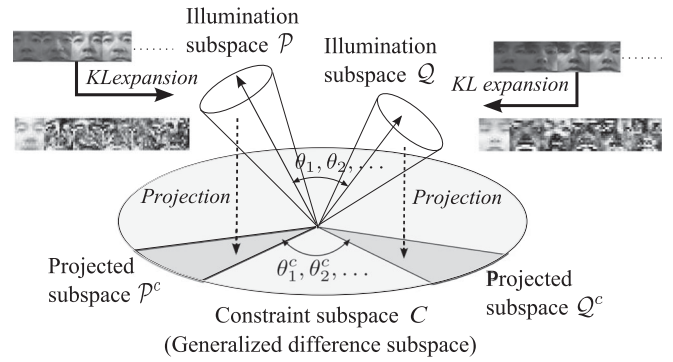

Fig. 1. Conceptual diagram of the constrained mutual subspace method. The canonical angles $\{\theta^C\}$ between the two projected subspaces $\mathcal{P}^C$ and $\mathcal{Q}^C$ onto the GDS are measured. Since it is not possible to depict the subspaces in a high-dimensional space, we show these schematically.

SM and MSM (CSM and CMSM), respectively. CSM and CMSM use the canonical angles θ^c between the projected subspaces, $\mathcal{P}^C$ and $\mathcal{Q}^C$, as a new measure of similarity. The term "constrained" reflects the fact that we perform conventional SM and MSM on a constraint subspace which satisfies some constraints. Various types of constraint subspaces are possible depending on the application. We choose to use GDSs as constraint subspaces in this paper.

Furthermore, we take nonlinearity of inputs into account. Although GDS projection works well for many data sets, the performance may degrade significantly as is the case with SM and MSM for image sets, such as multi-view images of 3D objects, due to their highly nonlinear structure. To cope with this problem, we use nonlinear subspaces generated by kernel principal component analysis (KPCA) [23] and then construct a nonlinear kernel DS (KDS) and a kernel GDS (KGDS) from these subspaces. We will show that KGDS can improve the performances of kernel SM [24], [25] and kernel MSM (KMSM) [11], [26] in the way that GDS can improve the performances of SM and MSM.

The main contributions of this paper are summarized as follows.

1) We describe the intrinsic characteristics of the types of difference subspaces (DS, GDS, KDS and KGDS) both theoretically and practically.

1-1) We confirm that these difference subspaces represent difference components between subspaces through their visualization by illumination subspaces of 3D objects.

1-2) We discuss a close relationship between projecting class subspaces onto a GDS/KGDS and the orthogonalization method of the class subspaces. We demonstrate that the projection works effectively as a pseudo-orthogonalization of class subspaces.

2) By using the above, we show that the optimal dimension of a GDS/KGDS should be set to the value that maximizes the overall mean of the canonical angles between projected class subspaces.

3) We demonstrate clearly by three different types of evaluation experiments that GDS/KGDS projection is an effective method to enhance the performance of conventional subspace-based methods.

Although some ideas in this paper have appeared in earlier publications [15], [27] in preliminary form, the present paper contains a more general formulation of our framework as well as the novel GDS, and provides a more in-depth discussion on the natures of the difference subspaces. We also report the results of new extensive experiments on object recognition, as well as on demonstrations of difference subspaces. These results are evidence for the utility of our proposed methods.

The rest of this paper is organized as follows. Section 2 reviews related work, especially the relationship to orthogonalization. Section 3 describes the concept of canonical angles and the algorithms of SM and MSM. In Sections 4 and 5, the novel concepts of DS and GDS are introduced and their optimal dimensionalities are discussed. KDS and KGDS are constructed in Section 6. In Section 7, GDS/KGDS projection is incorporated into MSM/KMSM. In Section 8, the characteristics of DS and GDS using synthetic and real 3D objects are shown, and the validity of GDS/KGDS projection is experimentally confirmed by using public and in-house databases. Section 9 summarizes our conclusions.

2 RELATED WORK

In this section, we describe the close relationship between GDS/KGDS projection in CSM and CMSM and the orthogonalization transformation (whitening transformation) [3], [22], a well-known method to enhance the performance of subspace-based methods such as SM and MSM [3], [28], [29], [30], [31].

The orthogonalization of given class subspaces can be accomplished by whitening the mixture auto-correlation matrix $\mathbf{R}$ of the basis vectors (the learning vectors) of all the class subspaces. Matrix $\mathbf{R}$ can be diagonalized by using an orthogonal matrix $\mathbf{U}$ as $\mathbf{U}\mathbf{R}\mathbf{U}^T = \Lambda$. The orthogonalization transformation matrix is then given as $\Lambda^{-1/2}\mathbf{U}$ [3], [22]. This equation means that the orthogonalization transformation has been realized by minimizing the variance of variances in the direction of basis vectors in the image vector space.

In contrast, projecting a subspace to GDS/KGDS is equivalent to removing the principal components of each class subspace, as described in Section 4. This operation discards basis vectors with larger variances and keeps those basis vectors with relatively smaller variances in the image vector space. Consequently, the variance of the variances due to the remaining basis vectors becomes smaller than that from the whole set of basis vectors. This means that characteristics similar to those achieved by the orthogonalization transformation are realized. Despite this positive effect, the above operation also has the drawback of reducing orthogonality between class subspaces by projecting a vector (resp. subspace) onto a GDS (resp. KGDS) with smaller dimension than the original space. Thus, these two factors need be considered to fully describe how GDS/KGDS projection achieves quasi-orthogonality. We opt to demonstrate the quasi-orthogonality by GDS/KGDS projection experimentally, as theoretical analysis is beyond the scope of this paper.

The orthogonalization method achieves the maximum possible orthogonality. In particular, complete orthogonalization can be achieved with a kernel mapping function.

However, performance degrades in cases where common subspaces exist between the given class subspaces. These common subspaces affect classification because orthogonalization does not remove them. In contrast, our method automatically removes common subspaces and is hence hardly affected by their existence.

In terms of orthogonalization, CMSM and kernel CMSM (KCMSM) are also related to the whitening MSM [29], the orthogonal subspace methods (OSM) [28], and the kernel orthogonal MSM (KOMSM) [31]. From what has been discussed so far, CMSM and KCMSM are expected to have the same level performance as those methods. Moreover, CMSM relates to discriminant analysis of canonical correlations (DCC) as discussed in [13]. DCC attempts to maximize the canonical angle of between-class subspaces while minimizing the canonical angles of within-class subspaces. This operation is closely related to orthogonalization.

In a broader sense, the proposed methods are also related to analysis on the Grassmann manifold [32], [33], [34]. It is defined by a set of r-dimensional subspaces in an f-dimensional vector space as $Gr(r, f)$. In this framework, each subspace is treated as a point on the Grassmann manifold. In this respect, MSM can be seen as the simplest form using the distance between two points on the manifold. Hence, for our reference, we also study the performance of Grassmann discriminate analysis (GDA) [32], an example of Grassmann manifold–based methods, and additionally discuss its relation to our proposed methods.

3 THE FORMULATION OF SUBSPACE-BASED METHODS

In this section we define canonical angles, and then outline the MSM algorithm.

3.1 Definition of Canonical Angles

Suppose an f-dimensional vector space with an N_p-dimensional subspace $\mathcal{P}$ and an N_q-dimensional subspace $\mathcal{Q}$. For convenience, we suppose $N_p \leq N_q$. The canonical angles $\{0 \leq \theta_1, \ldots, \theta_{N_p} \leq \frac{\pi}{2}\}$ between $\mathcal{P}$ and $\mathcal{Q}$ are recursively defined as follows [17], [18]:

$$\cos\theta_i = \max_{\mathbf{u} \in \mathcal{P}} \max_{\mathbf{v} \in \mathcal{Q}} \mathbf{u}^T\mathbf{v} = \mathbf{u}_i^T\mathbf{v}_i, \qquad (1)$$

$s.t. \|\mathbf{u}\| = \|\mathbf{v}\| = 1, \mathbf{u}_i^T\mathbf{u}_j = \mathbf{v}_i^T\mathbf{v}_j = 0, j = 1, \ldots, i-1,$

where $\mathbf{u}_i$ and $\mathbf{v}_i$ are the canonical vectors that form the ith smallest canonical angle, θ_i. The first canonical angle θ_1 is the smallest angle between $\mathcal{P}$ and $\mathcal{Q}$. The second smallest angle θ_2 is the smallest angle in a direction orthogonal to that of θ_1. The remaining θ_i for $i = 3, \ldots, N_p$ are calculated analogously, in a direction orthogonal to all smaller canonical angles.

There are several methods to calculate canonical angles [9], [17], [18]. The simplest and most practical method is singular value decomposition (SVD). Assume that $\mathbf{A} = [\boldsymbol{\Phi}_1 \ldots \boldsymbol{\Phi}_{N_p}] \in R^{f \times N_p}$ and $\mathbf{B} = [\boldsymbol{\Psi}_1 \ldots \boldsymbol{\Psi}_{N_q}] \in R^{f \times N_q}$ form unitary orthogonal bases for the two subspaces $\mathcal{P}$ and $\mathcal{Q}$. Let the SVD of $\mathbf{A}^T\mathbf{B} \in R^{N_p \times N_q}$ be $\mathbf{A}^T\mathbf{B} = \mathbf{U}\boldsymbol{\Sigma}\mathbf{V}^T, s.t. \boldsymbol{\Sigma} = \text{diag}(\kappa_1, \ldots, \kappa_{N_p})$. The canonical angles $\{\theta_i\}_{i=1}^M$ can be obtained as $\{cos^{-1}(\kappa_1), \ldots, cos^{-1}(\kappa_{N_p})\}$ $(\kappa_1 \geq, \ldots, \geq \kappa_{N_p})$.

The corresponding canonical vectors, $\mathbf{u}_i, \mathbf{v}_i (i = 1, \ldots, N_p)$ are obtained by the equations $[\mathbf{u}_1 \ldots \mathbf{u}_{N_p}] = \mathbf{AU}$ and $[\mathbf{v}_1 \ldots \mathbf{v}_{N_p}] = \mathbf{BV}$. The similarity between the two subspaces $\mathcal{P}$ and $\mathcal{Q}$ is measured by t angles as follows:

$$S[t] = \frac{1}{t} \sum_{i=1}^{t} \cos^2 \theta_i, \quad 1 \leq t \leq N_p. \tag{2}$$

3.2 Process Flow of Mutual Subspace Method

MSM proceeds as follows. Let us consider a set of L images of a class k object $(k = 1, \ldots, C)$ and denote the images as $f(= w \times h)$ dimensional vectors, $\{\mathbf{x}_l^k\}_{l=1}^{L}$, for images with $w \times h$ pixels.

In the training phase,

1) an $f \times f$ auto-correlation matrix is computed as $\mathbf{R}^k = \frac{1}{L} \sum_{l=1}^{L} \mathbf{x}_l^k \mathbf{x}_l^{k^T}$ from the given $\{\mathbf{x}_l^k\}_{l=1}^{L}$,
2) the orthonormal basis vectors of an N_d-dimensional subspace $\mathcal{P}_k$ of each class k are obtained as the eigenvectors corresponding to the N_d largest eigenvalues of $\mathbf{R}^k$, and
3) all of the class subspaces $\{\mathcal{P}_k\}_{k=1}^{C}$ are generated.

In the testing phase,

1) an N_{in}-dimensional input subspace, $\mathcal{I}$, is generated from a set of L' input images $\{\mathbf{x}_l^{in}\}_{l=1}^{L'}$ for an unknown object,
2) the similarities $\{S^k\}_{k=1}^{C}$ between $\mathcal{I}$ and all the class subspaces $\{\mathcal{P}_k\}_{k=1}^{C}$ are calculated using Eq. (2), and
3) $\mathcal{I}$ is classified into the class subspace with which it exhibits the highest similarity.

Note that in the SM [1], [3], $\mathcal{I}$ is replaced with an input vector for the testing phase.

4 CONCEPT OF DIFFERENCE SUBSPACES

In this section, we first define the DS between two subspaces geometrically. This definition is based on the canonical vectors that form the canonical angles between the subspaces. We then offer an equivalent analytic definition of the difference subspace by using the orthogonal projection matrices of the subspaces.

4.1 Geometrical Definition of Difference Subspace

The difference subspace is a natural extension of the difference vector between two vectors, $\mathbf{u}$ and $\mathbf{v}$, as shown in Fig. 2a.

We denote the difference subspace between the N_p-dimensional subspace $\mathcal{P}$ and the N_q-dimensional subspace $\mathcal{Q}$ as $\tilde{\mathcal{D}}_2$. The tilde is to distinguish the difference subspace from the *generalized* difference subspace between $\mathcal{P}$ and $\mathcal{Q}$, and the index "2" represents the number of classes to be classified. This will be defined shortly.

When there is no intersection between the two subspaces, N_p canonical angles $\{\theta_i\}_{i=1}^{N_p}$ (we assume $N_p \leq N_q$) are obtained between the subspaces as shown in Fig. 2b. Let $\mathbf{d}_i$ be the difference vector $\mathbf{u}_i - \mathbf{v}_i$ between canonical vectors $\mathbf{u}_i \in \mathcal{P}$ and $\mathbf{v}_i \in \mathcal{Q}$, which form the ith canonical angle θ_i. The vectors $\{\mathbf{d}_i\}_{i=1}^{N_p}$ are all mutually orthogonal, since the
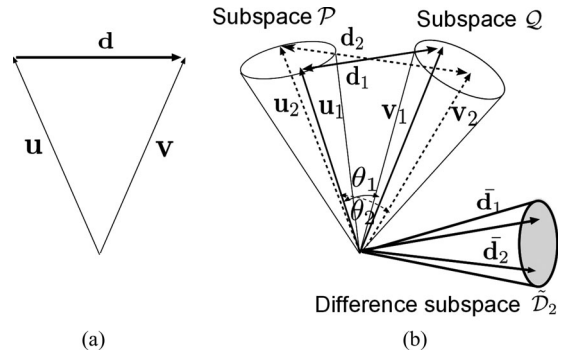

Fig. 2. Conceptual diagram of difference subspace: (a) difference vector $\mathbf{d}$ between $\mathbf{u}$ and $\mathbf{v}$; (b) difference subspace $\tilde{\mathcal{D}}_2$ between N_p-dimensional subspaces $\mathcal{P}$ and N_q-dimensional $\mathcal{Q}$ is defined with the canonical vectors $\mathbf{u}_i \in \mathcal{P}$ and $\mathbf{v}_i \in \mathcal{Q}$ forming the ith canonical angle θ_i between them. The orthogonal bases $\{\bar{\mathbf{d}}_i\}_{i=1}^{N_p}$ of $\tilde{\mathcal{D}}_2$ are obtained by normalizing the difference vectors $\mathbf{d}_i$ between $\mathbf{u}_i$ and $\mathbf{v}_i$.

directions of all the canonical angles are orthogonal to each other as described in Section 3. After normalizing the length of each $\mathbf{d}_i$ to 1, we regard the set of normalized difference vectors $\bar{\mathbf{d}}_i$ as an orthonormal basis of $\tilde{\mathcal{D}}_2$. $\tilde{\mathcal{D}}_2$ is then defined as $\langle \bar{\mathbf{d}}_1, \ldots, \bar{\mathbf{d}}_{N_p} \rangle$.

4.2 Analytic Definition of Difference Subspace

We now define $\tilde{\mathcal{D}}_2$ analytically by using the concept of an orthogonal projection matrix, which corresponds to the orthogonal projection operator onto a subspace. For the two subspaces $\mathcal{P}$ and $\mathcal{Q}$ considered in Section 3, the corresponding projection matrices $\mathbf{P}$ and $\mathbf{Q}$ are calculated as $\sum_{i=1}^{N_p} \Phi_i \Phi_i^T$ and $\sum_{i=1}^{N_q} \Psi_i \Psi_i^T$, respectively. In the following discussion, we assume that N_p is equal to N_q for simplicity.

Lemma. $\lambda_i = (\lambda_i^* - 1)^2$ where λ_i^* and λ_i are the ith largest eigenvalue of matrix $\mathbf{P} + \mathbf{Q}$ and that of matrix $\mathbf{PQ}$, respectively.

Proof. Let Δ^2 be a diagonal matrix whose entries are the positive eigenvalues of $\mathbf{P} + \mathbf{Q}$. Let $\mathbf{W}$ be a matrix with each column containing the normalized eigenvector of $\mathbf{P} + \mathbf{Q}$ for the corresponding eigenvalue in Δ^2. The relationship $(\mathbf{P} + \mathbf{Q})\mathbf{W} = \mathbf{W}\Delta^2$ holds. By premultiplying the above equation by $\mathbf{P}$ and $\mathbf{Q}$, we obtain the two equations $\mathbf{P}(\mathbf{QW}) = \mathbf{PW}(\Delta^2 - \mathbf{I})$ and $\mathbf{Q}(\mathbf{PW}) = \mathbf{QW}(\Delta^2 - \mathbf{I})$, respectively. Further, by substituting the second of these equations into the first, we obtain the equation $(\mathbf{PQ})(\mathbf{PW}) = (\mathbf{PW})(\Delta^2 - \mathbf{I})^2$. Hence, the relationship $\lambda = (\lambda_i^* - 1)^2$ holds. $\qquad \square$

Theorem. The ith basis vector $\bar{\mathbf{d}}_i$ of the difference subspace $\tilde{\mathcal{D}}_2$ is equal to the normalized eigenvector $\mathbf{x}_i$ of $\mathbf{P} + \mathbf{Q}$ that corresponds to the ith smallest eigenvalue smaller than 1.

Proof. From the Lemma, we can see that the eigenvalues λ_i^* of $\mathbf{P} + \mathbf{Q}$ are written with $1 \pm \sqrt{\lambda_i}$. Thus, we obtain the following equation for an eigenvector $\mathbf{x}$:

$$(\mathbf{P} + \mathbf{Q})\mathbf{x} = (1 \pm \sqrt{\lambda_i})\mathbf{x}. \tag{3}$$

Although the equation has two forms, depending on the sign of the right-hand side, we focus on the case where

the magnitude of the eigenvalue is smaller than 1.0. In this case, the eigenvalue of $\mathbf{P} + \mathbf{Q}$ is $1 - \sqrt{\lambda_i}$.

First, we consider the $i = 1$ case, as follows:

$$(\mathbf{P} + \mathbf{Q})\mathbf{x}_1 = (1 - \sqrt{\lambda_1})\mathbf{x}_1. \tag{4}$$

Note that $\mathbf{x}_1$ is the desired eigenvector, corresponding to the smallest eigenvalue of $\mathbf{P} + \mathbf{Q}$.

Since vector $\mathbf{x}_1$ is an eigenvector of $\mathbf{P} + \mathbf{Q}$, it can be represented by a linear combination of N_p pairs of canonical vectors, $\mathbf{u}_i \in \mathcal{P}$ and $\mathbf{v}_i \in \mathcal{Q}$, with weights a_i and b_i such that $\mathbf{x}_1 = \sum_{i=1}^{N_p} a_i \mathbf{u}_i + \sum_{i=1}^{N_p} b_i \mathbf{v}_i$. Substituting this equation into Eq. (4),

$$(\mathbf{P} + \mathbf{Q})\left(\sum_{i=1}^{N_p} a_i \mathbf{u}_i + \sum_{i=1}^{N_p} b_i \mathbf{v}_i\right)$$
$$= (1 - \sqrt{\lambda_1})\left(\sum_{i=1}^{N_p} a_i \mathbf{u}_i + \sum_{i=1}^{N_p} b_i \mathbf{v}_i\right). \tag{5}$$

The variables in the above equation have the following relationships for $\mu_i = \sqrt{\lambda_i}$: $\mathbf{P}\mathbf{v}_i = \mu_i \mathbf{u}_i$, $\mathbf{Q}\mathbf{u}_i = \mu_i \mathbf{v}_i$, $\mathbf{P}\mathbf{u}_i = \mathbf{u}_i$, and $\mathbf{Q}\mathbf{v}_i = \mathbf{v}_i$ [35]. Using these relationships, we obtain the following equation from Eq. (5).

$$\sum_{i=1}^{N_p} (b_i \mu_i + a_i \mu_1)\mathbf{u}_i + \sum_{i=1}^{N_p} (a_i \mu_i + b_i \mu_1)\mathbf{v}_i = 0. \tag{6}$$

Here, since the vectors $\{\mathbf{u}_1, \ldots, \mathbf{u}_{N_p}, \mathbf{v}_1, \ldots, \mathbf{v}_{N_p}\}$ are linearly independent, all the coefficients of the vectors $\mathbf{u}_i$ and $\mathbf{v}_i$ must be zero. Hence, the following relationships hold.

$$a_1 \mu_1 + b_1 \mu_1 = 0, \tag{7}$$

$$a_2 \mu_1 + b_2 \mu_2 = a_2 \mu_2 + b_2 \mu_1 = 0, \tag{8}$$

$$\vdots$$

$$a_{N_p} \mu_1 + b_{N_p} \mu_{N_p} = a_{N_p} \mu_{N_p} + b_{N_p} \mu_1 = 0. \tag{9}$$

From Eq. (7) we obtain that $a_1 = -b_1$, and from Eq. (8) that $(a_2 + b_2)(\mu_1 + \mu_2) = 0$. Here, since $\mu_1 + \mu_2 \neq 0$, also $a_2 = -b_2$. Thus, we obtain that $a_2(\mu_2 - \mu_1) = 0$ and $b_2(\mu_2 - \mu_1) = 0$ by substituting the above relationship into Eq. (8). Since $\mu_2 \neq \mu_1$, we obtain that $a_2 = b_2 = 0$. Similarly, the other coefficients, $\{a_i\}_{i=3}^{N_p}$ and $\{b_i\}_{i=3}^{N_p}$, are all zeros. Finally, the eigenvector $\mathbf{x}_1$, which satisfies Eq. (4), can be written as $\mathbf{x}_1 = a_1(\mathbf{u}_1 - \mathbf{v}_1)$.

In the same way, the eigenvector $\{\mathbf{x}_i\}_{i=2}^{N_p}$, which satisfies the equation $(\mathbf{P} + \mathbf{Q})\mathbf{x}_i = (1 - \sqrt{\lambda_i})\mathbf{x}_i$, can be written as $\mathbf{x}_i = a_i(\mathbf{u}_i - \mathbf{v}_i)$. Since $\mathbf{x}_i$ is a normalized eigenvector, it holds that $\bar{\mathbf{d}}_i = \mathbf{x}_i$ from the definition that $\mathbf{d}_i = \mathbf{u}_i - \mathbf{v}_i$. $\qquad \square$

Definition. *In addition to $\tilde{\mathcal{D}}_2$, We define the subspace that is spanned by the eigenvectors of $\mathbf{P} + \mathbf{Q}$ corresponding to eigenvalues larger than 1 as the principal component subspace (PCS), $\tilde{\mathcal{M}}_2$, in the sense that it consists of the principal components of the two subspaces $\mathcal{P}$ and $\mathcal{Q}$.*

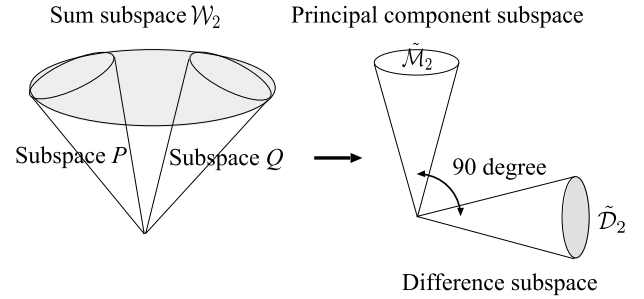The correspondence relationships described above can be summarized as follows.

Fig. 3. Direct sum decomposition of sum subspace $\mathcal{W}_2$ of $\mathcal{P}$ and $\mathcal{Q}$ into $\tilde{\mathcal{M}}_2$ and $\tilde{\mathcal{D}}_2$.

1) The N_p eigenvectors of matrix $\mathbf{P} + \mathbf{Q}$ that correspond to eigenvalues larger than 1 span the principal component subspace $\tilde{\mathcal{M}}_2$.
2) The N_p eigenvectors of matrix $\mathbf{P} + \mathbf{Q}$ that correspond to eigenvalues smaller than 1 span the difference subspace $\tilde{\mathcal{D}}_2$.

The relationships indicate that the sum subspace $\mathcal{W}_2$ of $\mathcal{P}$ and $\mathcal{Q}$, spanned by all the $N_p \times 2$ eigenvectors of matrix $\mathbf{P} + \mathbf{Q}$, can be directly decomposed to $\tilde{\mathcal{M}}_2$ and $\tilde{\mathcal{D}}_2$ such that $\mathcal{W}_2 = \tilde{\mathcal{M}}_2 \oplus \tilde{\mathcal{D}}_2$. Fig. 3 shows a conceptual diagram of this direct sum decomposition.

This decomposition suggests that the DS defined as the subspace representing the difference between two subspaces can be computed simultaneously with the subspace that is generated by removing the PCS of the two subspaces from their sum subspace.

Fig. 4 shows the three-dimensional basis vectors of subspaces $\tilde{\mathcal{M}}_2$ and $\tilde{\mathcal{D}}_2$ between two three-dimensional illumination subspaces of two synthetic hemispheres with identical centers and different diameters.

5 GENERALIZED DIFFERENCE SUBSPACE

Based on the analytic definition of difference subspace introduced in the preceding section, we generalize the concept of the difference subspace to the case of differences between two or more subspaces.

5.1 Generalization with Projection Matrices

Fig. 5 shows the conceptual diagram of the generalized difference subspace for $C \geq 2$ subspaces.

Given $C(\geq 2)$ N-dimensional class subspaces, $\mathcal{P}_k(k = 1, \ldots, C)$, a GDS $\mathcal{D}_C$ can be defined as the subspace
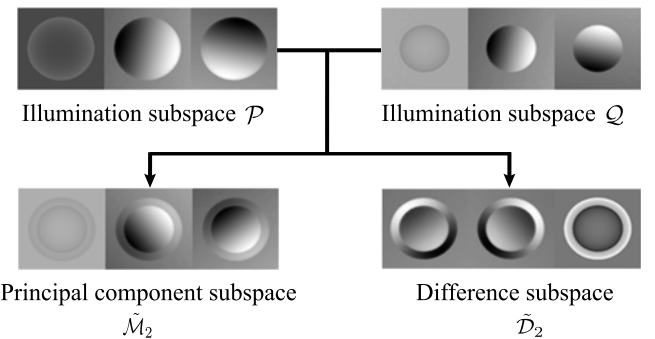

Fig. 4. Examples of $\tilde{\mathcal{M}}_2$ and $\tilde{\mathcal{D}}_2$ between two 3-dimensional illumination subspaces $\mathcal{P}$ and $\mathcal{Q}$.
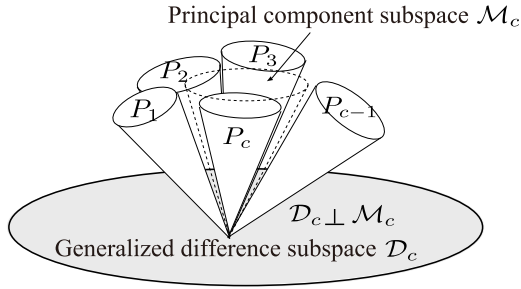
Fig. 5. Conceptual diagram of GDS $\mathcal{D}_C$ for C subspaces $\{\mathcal{P}_k\}_{k=1}^C$, obtained by removing the PCS $\mathcal{M}_C$ from the sum subspace of the subspaces.

Fig. 6. PCS $\mathcal{M}_3$ and GDS $\mathcal{D}_3$ of three three-dimensional illumination subspaces of different synthetic objects.

produced by removing the PCS $\mathcal{M}_C$ of all the class subspaces from the sum subspace $\mathcal{W}_C$ of those subspaces. From this definition, the GDS $\mathcal{D}_C$ is spanned by N_d eigenvectors, $\{\mathbf{d}_i\}_{i=N \times C - N_d + 1}^{N \times C}$ corresponding to the N_d smallest eigenvalues, λ_i, of the sum matrix $\mathbf{G} = \sum_{k=1}^{C} \mathbf{P}_k$ of orthogonal projection matrices $\mathbf{P}_k$ of the class k. Note that the GDS $\mathcal{D}_2$ coincides with $\tilde{\mathcal{D}}_2$ if the dimensions of $\mathcal{M}_2$ and $\mathcal{D}_2$ are equal. This equation may appear to be an equation of PCA in that basis vectors of all the subspaces are regarded as sample vectors. However, it differs from PCA because all sample vectors belonging to the same class should be orthogonal to each other. $\mathcal{D}_C$ contains only the essential component for discriminating all the classes, since it is orthogonal to the principal component subspace $\mathcal{M}_C$ which represents the principal component of all the class subspaces.

Fig. 6 shows the basis vectors of a six-dimensional $\mathcal{D}_3$ between the three-dimensional illumination subspaces of three planes with three small and one large hemispheres. The large hemispheres are put at the center, and the other three small hemispheres are put at each corner. The visualization result shows that the basis vectors of the $\mathcal{D}_3$ contain no information on the two common hemispheres (the small hemisphere at the upper left and the large hemisphere at the center) but retaining information of the other hemispheres as a difference.

Fig. 7 shows the orthonormal basis vectors of $\mathcal{M}_{128}$ and $\mathcal{D}_{128}$ of 128 16-dimensional illumination subspaces of the front faces of 128 subjects from the CMU Multi-PIE database [36]. This database consists of face images of 337 subjects, captured from 15 viewpoints with 20 lighting conditions in four recording sessions. In this visualization, we used front face images only of 128 subjects, with images collected across all four sessions. We took a subsampled image of size 16×16 pixels from an original image; we cropped this image by reference to the two inner corners of the eyes and the tip of the nose. The dimension of $\mathcal{D}_{128}$ was set to 211 according to a criterion that will be discussed next.

5.2 Selection of the Optimal Dimension of GDS

To consider how to select the optimal dimension of a GDS, we measured how the canonical angles between illumination subspaces are expanded by projecting them onto the GDS. As the measure, we introduced the orthogonal degree, $\omega = 1.0 - \bar{S}$, between two projected subspaces over all the pairs, where $\bar{S}$ is the mean of the similarities S in Eq. (2). As
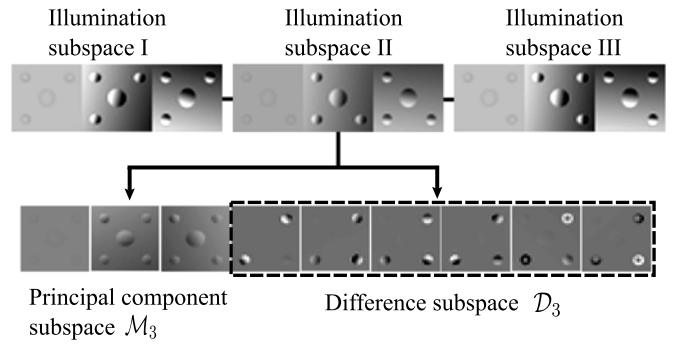
the relationship between the subspaces approaches perfect orthogonality, ω approaches 1.0.

In addition, we measured three performance indexes: class separability, error equal rate (EER), and recognition error rate (ER), to examine the extent to which the orthogonal degree corresponds to the discriminative ability of the GDS. Class separability represents the ratio of the similarities within classes and that between classes, and is equivalent to the Fisher criterion. The recognition performance is higher as the separability is closer to 1.0.

These indexes were measured by applying MSM to the projected testing subspaces while varying the dimension of the GDS from 256 to 60. The reference subspace of each subject was generated from 20 front face images in a session subset of the data set used in the preceding section, and the testing subspaces were generated from the subsets in the remaining three sessions.

The value of ω was also measured by using the three testing subsets as the reference subspaces. We repeated the above measurements four times by changing the combination of the reference and testing subspaces. Finally, we used the means of all the indexes as the final values of ω and as three performance indexes.

Fig. 8 shows changes in the resulting indexes against the dimension of the GDS. When the dimension is 256, all the indexes are the same as for MSM. Here note that ω was
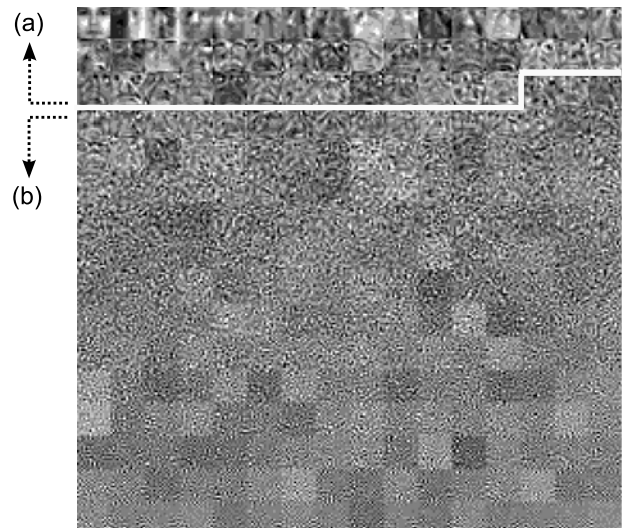

Fig. 7. Orthogonal basis vectors of $\mathcal{M}_{128}$ in the upper region (a) with 45 dimensions, and $\mathcal{D}_{128}$ in the remaining region (b) with 211 dimensions.
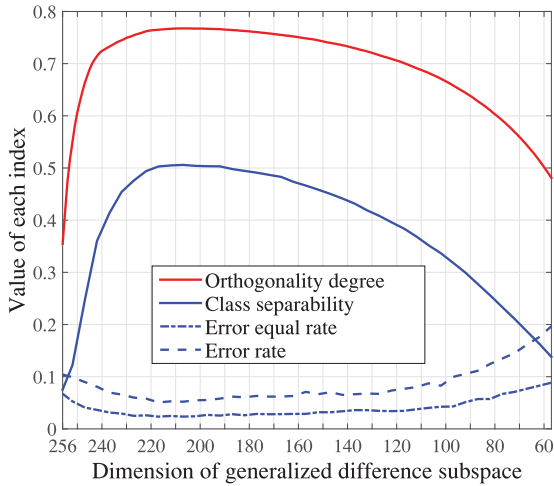
Fig. 8. Changes in the orthogonal degree between the projected subspaces and three performance indexes against the dimension of the GDS.

calculated as a predictive value by using just the reference subspaces in the learning phase. The three indexes were the actual values measured using both the reference and testing subspaces in testing phase.

The graph indicates clearly that the trend of the value of ω corresponds well to those of the three performance indexes. In particular, it is worth noting that all the indexes reach a maximum near 211 dimensions, at which ω has achieved a maximum value of 0.89 (ω is 0.31 without projection).

In the previous case (Fig. 6), ω achieved a maximum of 0.73 at dimension 6 of the GDS, at which dimension the differences among the three objects appeared most clearly. In the case of four three-dimensional subspaces of four 3D objects, ω achieved a maximum of 0.88 at dimension 9, at which dimension the differences among the objects were most clear.

These observations lead to an idea: we should select the dimension at which the value of ω is maximum. This criterion is easy to measure, and can be accurately estimated even from a subset of all the class reference subspaces.

6 NONLINEAR KERNEL GENERALIZED DIFFERENCE SUBSPACE

In this section, to deal with a set of multi-view images of a 3D object, we extend the GDS to nonlinear kernel GDS by using nonlinear kernel principal component analysis [23].

6.1 Generation of Nonlinear Class Subspaces

Let ϕ be the nonlinear function that maps the patterns $\mathbf{x} = (x_1, \ldots, x_f)^T$ of an f-dimensional input space $\mathcal{I}$ onto an f_ϕ-dimensional feature space $\mathcal{F}$, $\phi : R^f \to R^{f_\phi}, \mathbf{x} \to \phi(\mathbf{x})$, where $f_\phi \gg f$.

Given an N-dimensional nonlinear subspace $\mathcal{V}_k$ of class k generated by applying KPCA to L training patterns $\{\mathbf{x}_l^k\}_{l=1}^L$, the N orthonormal basis vectors $\{\mathbf{e}_i^k\}_{i=1}^N$, which span the nonlinear subspace $\mathcal{V}_k$, can be represented by the linear combination of $\{\phi(\mathbf{x}_l^k)\}_{l=1}^L$ as $\mathbf{e}_i^k = \sum_{l=1}^L a_{il}^k \phi(\mathbf{x}_l^k)$, where the coefficient a_{il}^k is the lth component of the eigenvector $\mathbf{a}_i^k$ corresponding to the ith largest eigenvalue λ_i of the $L \times L$ Gram matrix $\mathbf{K}$. The elements of matrix $\mathbf{K}$ are defined as

$[k_{ll'}]$, where $k_{ll'} = (\phi(\mathbf{x}_l^k) \cdot \phi(\mathbf{x}_{l'}^k)) = k(\mathbf{x}_l^k, \mathbf{x}_{l'}^k)$. We use an exponential kernel function $k(\mathbf{x}, \mathbf{y}) = \exp(-\frac{\|\mathbf{x} - \mathbf{y}\|^2}{\sigma^2})$. The vector of $\mathbf{a}_i^k$ is normalized to satisfy $\lambda_i(\mathbf{a}_i^k \cdot \mathbf{a}_i^k) = 1$.

6.2 Generation of KGDS

We generate N_d^ϕ-dimensional KGDS $\mathcal{D}^\phi$ from the set of C N-dimensional nonlinear class subspaces $\{\mathcal{V}_k\}_{k=1}^C$. The orthonormal basis of a KGDS can be obtained from all the orthonormal basis vectors of all the nonlinear subspaces; that is, the $C \times N$ basis vectors are used to find the orthonormal basis. This calculation is equivalent to applying PCA to all basis vectors. Let $\mathbf{E}$ be the matrix that contains all the basis vectors as columns: $\mathbf{E} = [\mathbf{e}_1^1 \ldots \mathbf{e}_N^1 \ldots \mathbf{e}_1^C \ldots \mathbf{e}_N^C]$. Then, we solve the eigenvalue problem of the matrix $\mathbf{D}$, defined as $\mathbf{E}^T \mathbf{E}$. In the calculation of each element of the matrix $\mathbf{D}$, the inner product between the ith orthonormal basis vector $\mathbf{e}_i^k$ of the subspace of class k and the jth orthonormal basis vector $\mathbf{e}_j^{k'}$ of the subspace of class k' can be obtained as the linear combination of kernel functions $k(\mathbf{x}_l^k, \mathbf{x}_{i'}^{k'})$ as follows:

$$(\mathbf{e}_i^k \cdot \mathbf{e}_j^{k'}) = \left(\sum_{l=1}^L a_{il}^k \phi(\mathbf{x}_l^k) \cdot \sum_{l'=1}^L a_{jl'}^{k'} \phi(\mathbf{x}_{l'}^{k'}) \right) \tag{10}$$

$$= \sum_{l=1}^L \sum_{l'=1}^L a_{il}^k a_{jl'}^{k'} (\phi(\mathbf{x}_l^k) \cdot \phi(\mathbf{x}_{i'}^{k'})) \tag{11}$$

$$= \sum_{l=1}^L \sum_{l'=1}^L a_{il}^k a_{jl'}^{k'} k(\mathbf{x}_l^k, \mathbf{x}_{l'}^{k'}). \tag{12}$$

The ith orthonormal basis vector $\mathbf{d}_i^\phi$ of the KGDS $\mathcal{D}^\phi$ can be represented as a linear combination of the vectors $\{\mathbf{E}_j\}_{j=1}^{C \times N}$, i.e. $\mathbf{d}_i^\phi = \sum_{j=1}^{C \times N} b_{ij} \mathbf{E}_j$ where $\mathbf{E}_j$ indicates the jth column of the matrix $\mathbf{E}$, and the weighting coefficient b_{ij} is the jth component of the eigenvector $\mathbf{b}_i$ that corresponds to the ith smallest eigenvalue β_i of the matrix $\mathbf{D}$ under the condition that the vector $\mathbf{b}_i$ is normalized to satisfy that $\beta_i(\mathbf{b}_i \cdot \mathbf{b}_i) = 1$.

$\mathbf{E}_j$ corresponds to the $\eta(j)$th basis vector of class $\zeta(j)$, where $\eta(j) = \mod(j-1, N) + 1$ and $\zeta(j) = \text{floor}(j-1/N) + 1$. The above notation for $\mathbf{d}_i^\phi$ may be written as follows:

$$\sum_{j=1}^{C \times N} b_{ij} \mathbf{E}_j = \sum_{j=1}^{C \times N} b_{ij} \sum_{l=1}^L a_{\eta(j)l}^{\zeta(j)} \phi(\mathbf{x}_l^{\zeta(j)}) \tag{13}$$

$$= \sum_{j=1}^{C \times N} \sum_{l=1}^L b_{ij} a_{\eta(j)l}^{\zeta(j)} \phi(\mathbf{x}_l^{\zeta(j)}). \tag{14}$$

6.3 Projection onto KDS

It is impossible to visualize the orthonormal basis vectors $\mathbf{d}_i^\phi$ of the KGDS $\mathcal{D}^\phi$. However, the projection $\tau(\phi(\mathbf{x}))$ of the mapped pattern $\phi(\mathbf{x})$ onto the basis vectors $\mathbf{d}_i^\phi$ can be calculated from an input pattern $\mathbf{x}$ and all $L \times C$ training patterns $\{\mathbf{x}_l^k\}_{l=1}^L (k = 1, \ldots, C)$ as follows:

$$(\phi(\mathbf{x}) \cdot \mathbf{d}_i^\phi) = \sum_{j=1}^{C \times N} \sum_{l=1}^L b_{ij} a_{\eta(j)l}^{\zeta(j)} (\phi(\mathbf{x}_l^{\zeta(j)}) \cdot \phi(\mathbf{x})) \tag{15}$$

$$= \sum_{j=1}^{C \times N} \sum_{l=1}^L b_{ij} a_{\eta(j)l}^{\zeta(j)} k(\mathbf{x}_l^{\zeta(j)}, \mathbf{x}), \tag{16}$$

where we can easily compute $k(\mathbf{x}_l^{\zeta(j)}, \mathbf{x})$ through $k(\mathbf{x}, \mathbf{y}) = \exp(-\frac{\|\mathbf{x}-\mathbf{y}\|^2}{\sigma^2})$. Finally, the projection $\tau(\phi(\mathbf{x}))$ of the mapped $\phi(\mathbf{x})$ onto the $N_d^\phi(< C \times N)$ dimensional KGDS is represented as following:

$$\tau(\phi(\mathbf{x})) = (z_1, \ldots, z_{N_d^\phi})^T, z_i = (\phi(\mathbf{x}) \cdot \mathbf{d}_i^\phi). \quad (17)$$

7 INCORPORATION OF GDS/KGDS PROJECTION INTO MSM/KMSM

In this section, we describe the CMSM and KCMSM algorithms. As described in Section 1, the essence of CMSM/KCMSM is to conduct MSM/KMSM on GDS/KGDS. Thus, the key matter of interest here is the way to project the input subspace and class subspaces onto the GDS/KGDS. There are two ways that give equivalent results. One is to project the basis vectors of each subspace onto GDS/KGDS and then normalize the projected basis, which is further followed by Gram-Schmidt orthogonalization after the orthonormality is lost by the projection. The alternative is to first project the sample images for each subspace and then generate a subspace from the projected images. We base our description on this second method as it is more straightforward to understand.

7.1 Algorithm of the CMSM

Given: a set of C N-dimensional class subspaces $\{\mathcal{P}_k\}_{k=1}^C$ generated from the set of the L training image patterns $\{\mathbf{x}_l^k\}_{l=1}^L$ of each class k.

In the training phase,

1) an N_d-dimensional GDS $\mathcal{D}_c$ is generated from a set of the class subspaces $\{\mathcal{P}_k\}_{k=1}^C$ in the way described in Section 5,
2) the training images $\{\mathbf{x}_l^k\}_{l=1}^L$ of class k are projected onto the $\mathcal{D}_c$, and
3) the N-dimensional class subspace $\mathcal{P}_k^\mathcal{D}$ of class k is generated by applying PCA to the projected images $\{\tau(\mathbf{x}_l^k)\}_{l=1}^L$.

In the testing phase,

1) an N_{in}-dimensional input subspace $\mathcal{P}_{in}^\mathcal{D}$ is generated from the set of the projected input images $\{\tau(\mathbf{x}_{l'}^{in})\}_{i'=1}^{L'}$ onto the $\mathcal{D}_c$ (note that $\mathcal{P}_{in}^\mathcal{D}$ is replaced by a vector for CSM),
2) the similarities $\{S^k\}_{k=1}^C$ between P_{in}^D and all the class subspaces $\{P_k^D\}_{k=1}^C$ are calculated by using Eq. (2), and
3) the class subspace with the highest similarity is determined to be that of the identified class if the similarity exceeds a threshold.

7.2 Algorithm of the KCMSM

Fig. 9 shows the steps of the algorithm of KCMSM, exemplified by a hand shape classification using multiple view images.

Given: a set of C nonlinear class subspaces $\{\mathcal{P}_k\}_{k=1}^C$, obtained by applying KPCA to the set of the training images $\{\mathbf{x}_l^k\}_{l=1}^L$ of class $k(= 1, \ldots, C)$.
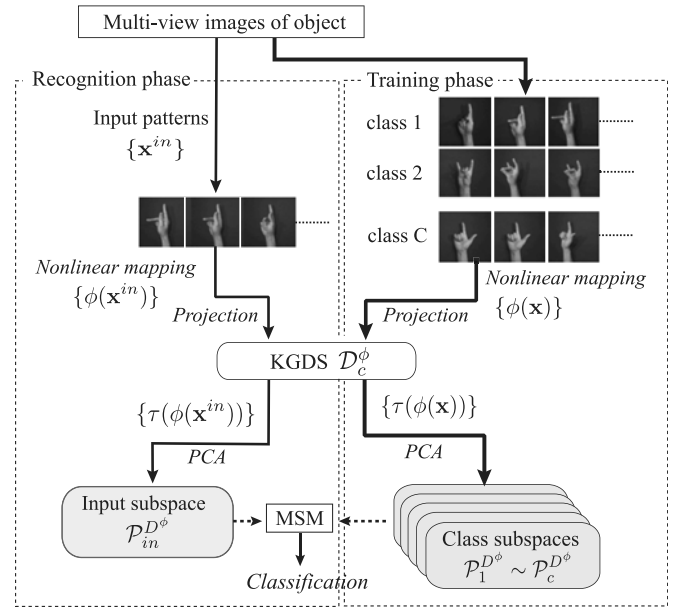

Fig. 9. The flow of KCMSM in hand shape classification.

In the training phase,

1) an N_d^ϕ-dimensional KGDS $\mathcal{D}_c^\phi$ is generated from the set of nonlinear class subspaces $\{\mathcal{P}_k\}_{k=1}^C$ in the way described in Section 6,
2) the nonlinear mapped images $\{\phi(\mathbf{x}_l^k)\}$ of the images $\{\mathbf{x}_l^k\}$ are projected onto the $\mathcal{D}_c^\phi$ by using Eq. (17), and
3) an N-dimensional linear class subspace $\mathcal{P}_k^{D^\phi}$ of each class k is generated by applying standard PCA to the set of the projected images $\{\tau(\phi(\mathbf{x}_l^k))\}$.

In the testing phase,

1) the nonlinear mapped images $\{\phi(\mathbf{x}_{l'}^{in})\}$ of a set of input images $\{\mathbf{x}_{i'}^{in}\}_{l'=1}^{L'}$ are projected onto the KGDS $\mathcal{D}_c^\phi$,
2) an N_{in}-dimensional linear input subspace $\mathcal{P}_{in}^{D^\phi}$ is generated by applying standard PCA to a set of the projected images $\{\tau(\phi(\mathbf{x}_{l'}^{in}))\}$,
3) the similarities $\{S^k\}_{k=1}^C$ between the input subspace $\mathcal{P}_{in}^{D^\phi}$ and all the class subspaces $\{\mathcal{P}_k^{D^\phi}\}_{k=1}^C$ are calculated by Eq. (2), and
4) the input subspace is classified into the class subspace with the highest similarity S if the similarity exceeds a threshold.

8 EVALUATION EXPERIMENTS

We first validate that the various difference subspaces are equivalent to the difference between multiple subspaces through visualization of several difference subspaces for multiple Lambertian objects with uniform albedo. The images used here are all real. In this validation we considered the simplest situation, where the relative position between camera and object was fixed. Next, we confirm the validity of GDS/KGDS projection for the subspace-based methods.

8.1 Difference Extraction with DS and GDS

To examine the characteristic of DS, we visualized the orthogonal basis vectors of the DS between illumination

(a) Sample images of cow relief-1 with a scoop.

(b) Sample images of cow relief-2.

(c) Orthonormal basis vectors spanning the PCS (upper) and the DS (lower).

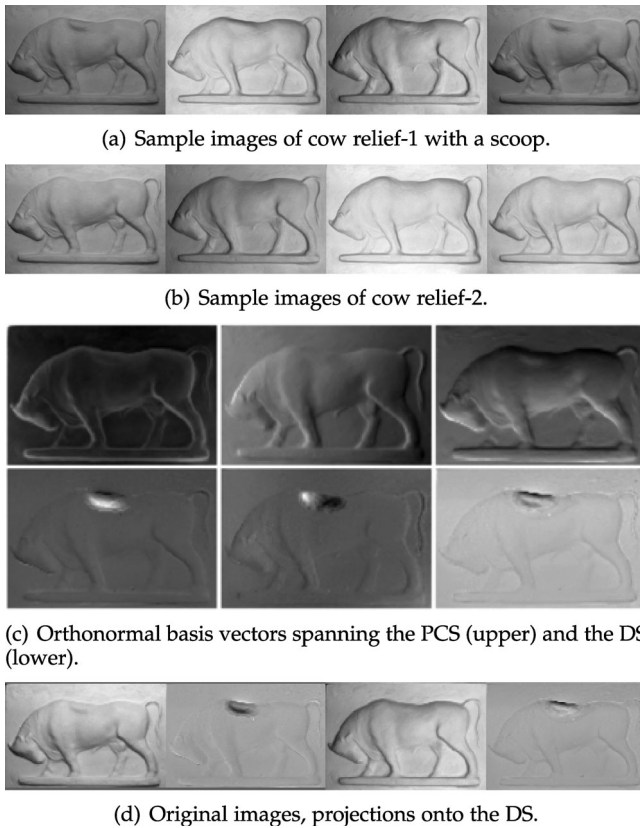(d) Original images, projections onto the DS.

Fig. 10. PCS and DS of two real cow reliefs.

subspaces of two objects with similar overall shapes in the two following cases.

In the first case, we considered two complicated cow reliefs with a height of 20 cm and a width of 30 cm. The cow reliefs are slightly different in that relief-1 has a small scoop on its shoulder region, while relief-2 has no such scoop. Figs. 10a and 10b show examples of 420×640 pixel images of both the reliefs, which were captured from a fixed viewpoint with various illuminations. Their surfaces were coated by white paper clay and had no shadows, so that the surfaces can be almost regarded as Lambertian surfaces.

Fig. 10c shows the orthonormal basis vectors of the three-dimensional PCS and three-dimensional DS between two three-dimensional illumination subspaces of the reliefs. We can see that the scoop region is emphasized in the basis vectors of the DS. Fig. 10d shows the original images and the corresponding projected images, which demonstrate that the DS projection can selectively extract the region corresponding to the shape difference.

In the second case, we considered three balls of the same radius but with a different number of convex parts on their surfaces, as shown in Fig. 11a. Ball-1, Ball-2, and Ball-3 have two, three, and four convex parts, respectively. Their surfaces were coated by white paper clay. Thirty images of each ball were captured from a fixed camera view under various illumination conditions. A six-dimensional GDS was generated from the three three-dimensional illumination subspaces of the three balls. Fig. 11b shows the orthonormal basis of the three-dimensional PCS and the six-dimensional GDS. Only the two convex parts corresponding to the differences
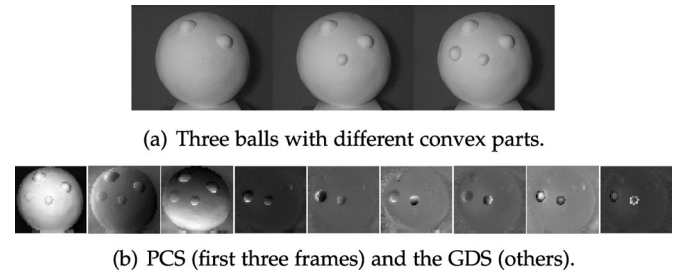

(a) Three balls with different convex parts.

(b) PCS (first three frames) and the GDS (others).

Fig. 11. PCS and GDS of three real balls.

among them are selectively emphasized in the basis vectors of the GDS.

8.2 Difference Extraction with KDS and KGDS

We visualized the orthonormal bases of KDS and KGDS in two cases.

In the first case, we considered two sets of multi-view images of two Venus sculptures, Venus-1 and Venus-2. They are slightly different: Venus-1 wears two earrings, but Venus-2 does not. Figs. 12a and 12b show examples of the multi-view images with size 63×48 pixels of the Venus sculptures rotating on a turntable under a fixed view and illumination condition. A 100-dimensional nonlinear subspace was generated from a set of the multi-view images of each Venus sculpture. A 100-dimensional KDS was generated from the two 100-dimensional subspaces. An exponential kernel function with $\sigma^2 = 5$ was used for the nonlinear mapping.

Since it is impossible to visualize an image projected onto a KDS/KGDS in the feature space due to its infinite

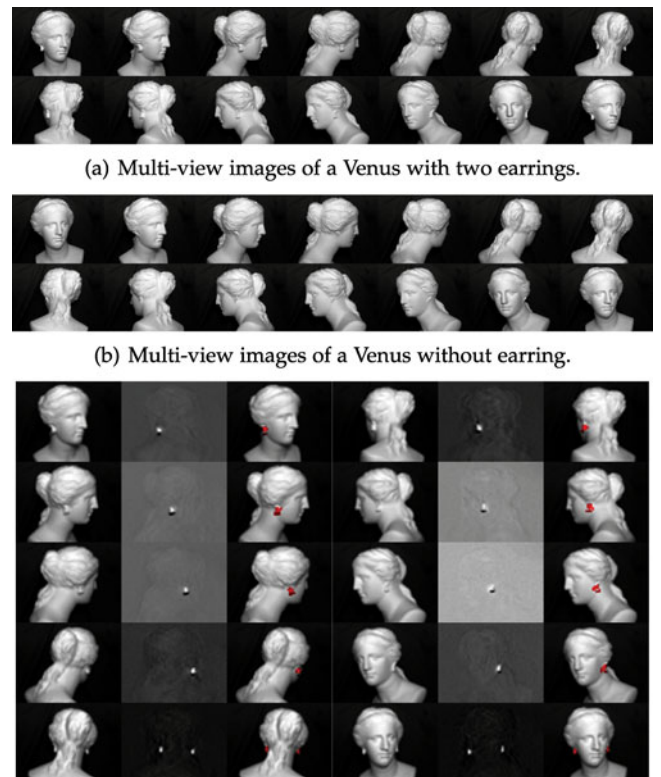(a) Multi-view images of a Venus with two earrings.

(b) Multi-view images of a Venus without earring.

(c) Input image, preimages corresponding to projected images onto the KDS, emphasized images.

Fig. 12. Projecting images onto nonlinear KDS.

(a) Venus-1 with a necklace and two earrings.

(b) Venus-2 with earrings.

(c) Venus-3 with no decoration.

(d) Input images, preimages of the projected images onto the KGDS, emphasized images.
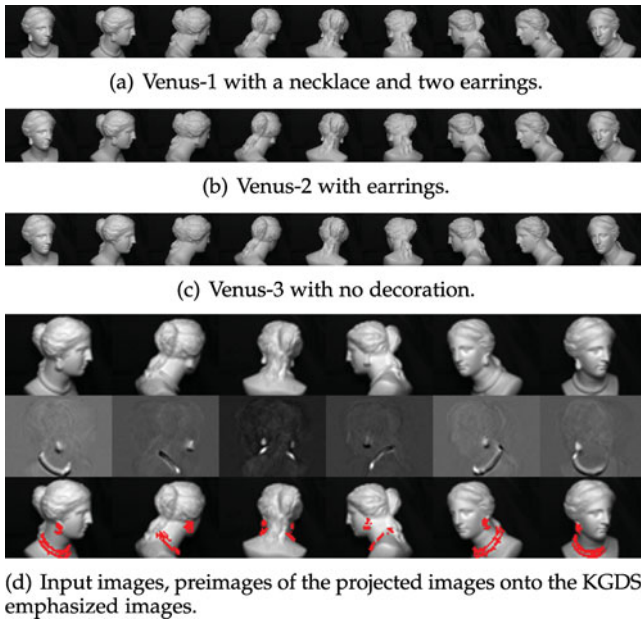
Fig. 13. Projecting images onto nonlinear KGDS.

dimensionality, we searched for the preimage in the original input space I, corresponding to the projected image, by using the search method [37]. Fig. 12c shows input images, the preimages corresponding to the projected input images, and the emphasized images of the input image from the preimages. The visualization results show that only the regions corresponding to the differing parts (the two earrings) are clearly extracted in the preimages; this is what was expected.

In the second case, we considered three kinds of Venus sculptures with slight differences. Venus-1 has a necklace and two earrings; Venus-2 wears two earrings; Venus-3 has no decorations. Three 150-dimensional nonlinear subspaces were generated from the set of multiview images of each Venus, and then a 300-dimensional KGDS was generated from the three 150-dimensional nonlinear subspaces. Fig. 13 shows input images, the preimages corresponding to the projected input images, and emphasized images. The results indicate clearly that the differences between the three Venuses, that is, only the regions of the two earrings and the necklace, were selectively extracted. These results provide evidence that the KDS/KGDS projection works well to extract the difference between multiple subspaces.

8.3 Validity of Projecting onto GDS for SM

In order to confirm the validity of GDS projection for the SM, we performed a facial recognition experiment with data from the widely used public database, Yale face database B+ [6].

In this experiment, we used the SM with a specific nine-dimensional illumination subspace as a baseline. This specific illumination subspace is generated from a set of nine front face images of a person under the nine specific lighting conditions shown in Fig. 14. As a result, the illumination subspace can accurately approximate an illumination cone, which consists of all the possible images of a face under various lighting conditions [7].

Fig. 14. Nine images of a person illuminated by nine specific lighting points.

We refer to this method as SM-$9PL$ method in this paper, to emphasize that it is based on SM (it is called the "nine points of lights method" in [7]). The experimental results in [7] give evidence that the SM-9PL significantly outperforms the conventional correlation, Eigenfaces [19], nearest neighbor, cones-attached [6], harmonic images [8], harmonic images-cast [8], and gradient angle [38] methods. The SM-9PL achieved performance comparable to the state-of-the-art cones-cast method [6]. The SM-9PL that incorporates the GDS projection is referred to as CSM-$9PL$. We compared the performances of SM-9PL with those of CSM-9PL.

8.3.1 Experimental Settings

The Yale face database B+ consists of face images of 38 subjects. These images were acquired under 64 different lighting conditions in nine different poses [7]. The database has four subsets, clustered according to the angle that the light source direction forms with the camera's axis. We selected 29 individuals from the database; these individuals' images were completely acquired across the four subsets. Only the front face images were used in this experiment, so that our evaluation database contains 1,305 images of 29 subjects under 45 different lighting conditions in the front pose. We converted the cropped images of 640×480 pixels to images of 32×24 pixels.

The class subspace of each subject was generated from the nine specific images used in the SM-9PL. A GDS was generated from the 29 nine-dimensional class subspaces. We set the dimension of the GDS to 231, at which the orthogonal degree, ω, was maximal. The remaining 36 images, corresponding to 36 lighting conditions except for the nine specific lighting conditions, were used as input images in the testing phase. The total number of trials was 1,044.

8.3.2 Experimental Results and Discussions

Table 1 shows a comparison between the performances of the SM-9PL and the CSM-9PL in terms of the EER. Note that the recognition rates of both were 100 percent due to using 9PL illumination subspaces, as discussed in [6], [7]. The GDS projection improves significantly the EER of the SM-9PL, from 2.73 to 0.07 percent. This advantage of the CSM-9PL can be also clearly seen in the receiver operating characteristic (ROC) curves as shown in Fig. 15.

8.4 Validity of Projecting onto GDS for MSM

To confirm the validity of GDS projection for MSM, we compared the performance of CSM/CMSM (GDS projection +

TABLE 1
EER(%) of the CSM-9PL and the SM-9PL

	Subset1	Subset2	Subset3	Subset4	Mean
SM-9PL	1.97	0.29	1.08	7.59	**2.73**
CSM-9PL	0	0	0.29	0.01	**0.07**

Fig. 15. ROC curves of the CSM-9PL and the SM-9PL.

(a) 30 classes of hand shapes.

(b) Seven-view images of hand #30.

Fig. 16. Testing data set for hand shape recognition.

SM/MSM) with that of SM/MSM and several related methods: OMSM (Orthogonalization + MSM) [29], [39], discriminant analysis of canonical correlations [13], sparse representation classification (SRC) [40], and Grassmann discriminate analysis [32]. GDA is the Fisher discriminate analysis (FDA) on a Grassmann manifold, $Gr(r, f)$ with $r = 20$ and $f = 256$ in this case, with the kernel of projection metric. We consider facial recognition from the CMU Multi-PIE database [36] under the same settings as in the evaluation in Section 5.

GDS projection and Fisher discriminative analysis (FDA) share the same purpose of effective extraction of discriminative features, although their mechanisms are different from each other. To extend our comparisons, we also evaluated the performance by the combination of FDA + MSM, in which the FDA was used as a pre-processing for the MSM.

8.4.1 Experimental Settings

SM and SRC use a single input vector whereas MSM requires multiple input images. To evaluate their performance under the same condition as in MSM, we use the mean of the similarity values of multiple input vectors as the similarity for the images. The dimensions of the class subspace and testing subspaces were fixed to 10 and 16, for DCC and the other methods, respectively. We generated the GDS from 128 class subspaces with dimension 16. The dimension of the GDS was set to 211, which corresponds to the maximal orthogonal degree as shown in Section 5.
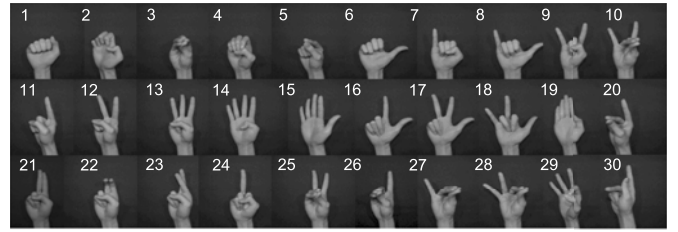
8.4.2 Experimental Results and Discussions

Table 2 shows the performance of all the considered methods in terms of ER (%) and EER (%). GDS projection significantly improves the performance of MSM. CMSM improves on CSM, which previously showed superior performance to that of conventional methods in earlier experiments. As expected, the rates achieved by CMSM are very similar to those of OMSM, which reflects that GDS projection is equivalent to orthogonalization. FDA did not improve

significantly on MSM, possibly because FDA extracted a feature optimal for methods based on Euclidean distance, but not necessarily for subspace-based methods such as SM and MSM.

The table shows that SRC and GDA also give good results. In particular, the low EER by GDA supports that FDA on the Grassmann manifold worked effectively. DCC did not perform as well as CMSM and GDA, which could be due to the comparatively few (20) training patterns in this experiment.

8.5 Validity of Projecting onto KGDS for KMSM

To verify the validity of KGDS projection for KMSM, we compared the KCMSM (KGDS projection + KMSM) with those of KMSM [11], [26] on the classification of 30 kinds of hand shapes. To confirm that KGDS projection has the same effect as the standard orthogonality method, we also compared the performances of KCMSM to those of kernel orthogonal MSM (KOMSM) [31]. As related subspace-based methods,[1] we evaluated the performance of GDA.

8.5.1 Experimental Settings

We collected the multi-view images of 30 kinds of hand shapes at 1 fps from 100 subjects for 4 seconds by using a multi-camera system equipped with seven synchronized cameras intervals of 10 degree [14]. During the collection, the subjects were asked to rotate their hands at a constant speed to increase the number of viewpoints. Fig. 16a shows the 30 kinds of hand shapes. Fig. 16b shows the seven-view images of hand #30.

A hand region was roughly cropped from an input image, resized to 25×25 pixels and vectorized to a 625-dimensional raw image vector. Stable segmentation of a hand region is difficult due to the different shape of every hand class and the difficulty in defining stable and reproducible feature points. To deal with this, we also used 313-dimensional FFT features as position invariant features. These were extracted by applying the fast Fourier transform

TABLE 2
The Performance of Different Methods on the CMU Database

	SM [1]	GDS +SM	DCC [13]	SRC [40]	GDA [32]	MSM [9]	FDA +MSM	Orth+ MSM [29]	GDS +MSM
ER	16.5	11.5	8.9	9.4	7.1	13.0	9.4	5.1	5.5
EER	13.3	5.7	6.5	2.5	1.4	7.8	4.1	2.4	2.4

---

1. For an option of applying Kernel FDA to KMSM as in the previous experiment, appropriate evaluation was not available for KSM/KMSM with KFDA because the dimension of the discriminate space generated by KFDA (=29) was too low for them to be applied.

TABLE 3
Experimental Results of Hand Shape Classification

	Raw image feature					
	MSM [9]	**GDS +MSM**	GDA [32]	KMSM [11], [26]	KOrth+ KMSM [31]	**KGDS +KMSM**
Test dim	5	5	20	20	20	20
Class dim	30	30	20	200	200	200
ER (%)	22.55	**17.12**	10.50	12.52	9.43	**9.28**
EER (%)	18.49	**13.77**	3.66	15.95	5.59	**5.15**
	FFT feature					
	MSM [9]	**GDS +MSM**	GDA [32]	KMSM [11], [26]	KOrth+ KMSM [31]	**KGDS +KMSM**
Test dim	5	5	20	7	7	7
Class dim	30	30	20	100	100	100
ER (%)	11.99	**8.78**	4.45	7.13	5.66	**5.84**
EER (%)	17.69	**13.00**	1.67	13.17	4.00	**3.29**

to the cropped images and then vectorizing half of the obtained frequency images. All the feature vectors were normalized in length.

To form training and testing data sets, we randomly divided subjects into two groups. The images of one group were used for training and those of the other were used for testing. We evaluated each of the algorithms on the data sets from ten random group divisions. The average of the ten trials was used as the final algorithm evaluation index. A nonlinear class subspace of each hand shape was generated from the set of 1,400 (7 cameras ×4 images ×50 subjects) images.

The testing subspaces were generated from 28 images of every combination of subject and shape. The total number of generated testing subspaces was 1,500 (=30 shapes ×50 subjects). The dimensions of the testing and class subspaces were set as shown in Table 3. KGDS was generated from the 200-dimensional 30 nonlinear class subspaces for Raw feature and its dimension was set to 4,200, at which the value of ω is maximal. For FFT feature, a 2,200-dimensional KGDS was generated from the 100-dimensional thirty nonlinear class subspaces in the same way. The values of σ^2 of kernel function was set to 0.8 and 0.145 for Raw and FFT features, respectively. For GDA Grassmann manifolds, $Gr(20, 625)$ and $Gr(20, 313)$ were utilized for Raw and FFT features, respectively. A 620-dimensional GDS for CMSM with Raw feature was generated from the thirty 30-dimensional linear class subspaces. A 311-dimensional GDS was also generated for FFT feature in the same way.

8.5.2 Experimental Results and Discussions

Table 3 shows the performance of all the methods in terms of ER and EER. The results clearly show that KGDS projection enhanced the performance of KMSM. The improvement was further amplified by using FFT feature images with position invariance. This suggests that we select effective position-invariant feature more carefully, although the subspace-based methods could achieve comparative practical performance using multiple images. The equality between KCMSM and KOMSM in the performance levels indicates that KGDS projection plays an equivalent role to orthogonalization.

8.5.3 Discussions in Relation to GDA

In terms of Grassmann manifold, MSM can be regarded as a simple 1-NN method using the distance between two points in the manifold where a DS is corresponding to a tangent vector at the limit. These interpretations further suggest that the projection on a GDS/KGDS may be well related to concepts such as discriminant analysis and tangent spaces on a Grassmann manifold. In our experiments the performance of GDA turned out to be equivalent or superior to that of our proposed method. This is likely because GDA exploits information of the variance within each class whereas KCMSM simply assumes homogeneity of variance, and because each class is represented by a set of multiple subspaces in GDA while it is represented by a single subspace in our method. Thus, it could well be the case that GDA outperforms our method for inputs such as coming from hand shape class that involve complicated distributions that are hard to capture by a single subspace.

On the other hand, unlike our methods, it is required in GDA that the dimensions of input and reference subspaces must be equal even in the case that the distributions of testing and learning data sets are unbalanced, and this could potentially pose a practical limitation. For example, let us assume the number of available images to be limited in the experiment of facial recognition (Section 8.4). In the case that we have seven input frames for instance, it is observed that the result of the GDS+MSM is better than GDA after all. That is, the ER and EER of GDA with $r = 7$ are 12.96 and 2.93 percent, respectively, while those of the GDS+MSM are 6.51 and 3.44 percent in the case of commonly using the same dimension, $N_{in} = N = 7$, but further down to 5.86 and 2.86 percent, respectively, in the original setting with $N = 16$ for the class subspace.

Further detailed analysis on the relationship to GDA is beyond the scope of this paper, but we note that our methods are simple and therefore widely applicable, thanks to the general mathematical concept of the difference subspace.

9 CONCLUSIONS

This paper proposed a framework for extracting differences between multiple subspaces to fundamentally address the object recognition problem by subspace-based methods. The framework was built on a new concepts of DS and a generalization of this idea, GDS. First, DS was defined as a subspace corresponding to the difference between two subspaces, a natural extension of a difference vector between two vectors. Then, DS was generalized to GDS to allow more than two subspaces. DS and GDS were then extended to nonlinear kernel versions, KDS and KGDS. By visualizing the difference subspaces between illumination subspaces of 3D objects, we confirmed that the various types of difference subspaces indicate the difference between multiple subspaces.

We discussed and demonstrated a close relationship between GDS/KGDS projection and Fukunaga and Koontz's orthogonalization method. This relationship led to the idea that the optimal dimensions of a GDS/KGDS should be set such that the orthogonal degree among projected class subspaces is maximized. We incorporated GDS/KGDS projection into the family of subspace-based methods to enhance their performances. The validity of the proposed framework

was extensively evaluated through 3D object recognition experiments with data from the Yale face database B+, the CMU Multi-PIE database, and a hand shape database created for these experiments. GDS/KGDS projection are also related to analysis on the Grassmann manifold. Thus, we discussed their relation through evaluations of the performance of Grassmann discriminate analysis.

ACKNOWLEDGMENTS

REFERENCES

[1]   S. Watanabe and N. Pakvasa, "Subspace method of pattern recognition," in *Proc. 1st Int. Conf. Pattern Recognit.*, 1973, pp. 25–32.

[2]   T. Iijima, H. Genchi, and K. Mori, "A theory of character recognition by pattern matching method," in *Proc. 1st Int. Conf. Pattern Recognit.*, 1973, pp. 50–56.

[3]   E. Oja, *Subspace Methods of Pattern Recognition*. Research Studies Press Ltd., Letchworth, England, 1983.

[4]   A. Shashua, "On photometric issues in 3d visual recognition from a single 2d image," *Int. J. Comput. Vis.*, vol. 21, pp. 99–122, 1997.

[5]   P. N. Belhumeur and D. J. Kriegman, "What is the set of images of an object under all possible lighting conditions?" *Int. J. Comput. Vis.*, vol. 28, pp. 1–16, 1998.

[6]   A. S. Georghiades, P. N. Belhumeur, and D. J. Kriegman, "From few to many: Illumination cone models for face recognition under variable lighting and pose," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 23, no. 6, pp. 643–660, Jun. 2001.

[7]   K. C. Lee, J. Ho, and D. J. Kriegman, "Acquiring linear subspaces for face recognition under variable lighting," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 27, no. 5, pp. 684–698, May 2005.

[8]   R. Basri and D. W. Jacobs, "Lambertian reflectance and linear subspaces," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 25, no. 5, pp. 218–233, May 2003.

[9]   K. Maeda and S. Watanabe, "A pattern matching method with local structure," *Trans. IEICE*, vol. J68-D, pp. 345–352, 1985.

[10]   O. Yamaguchi, K. Fukui, and K. Maeda, "Face recognition using temporal image sequence," in *Proc. Int. Conf. Autom. Face Gesture Recognit.*, 1998, pp. 318–323.

[11]   H. Sakano and N. Mukawa, "Kernel mutual subspace method for robust facial image recognition," in *Proc. Int. Conf. Knowl.-Based Intell. Eng. Syst. Allied Technol.*, 2000, vol. 1, pp. 245–248.

[12]   A. Maki and K. Fukui, "Ship identification in sequential ISAR imagery," *Mach. Vis. Appl.*, vol. 15, no. 3, pp. 149–155, 2004.

[13]   T. K. Kim, J. Kittler, and R. Cipolla, "Discriminative learning and recognition of image set classes using canonical correlations," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 29, no. 6, pp. 1005–1018, Jun. 2007.

[14]   Y. Ohkawa and K. Fukui, "Hand shape recognition using the distributions of multi-viewpoint image sets," *IEICE Trans. Inform. Syst.*, vol. E95-D, no. 6, pp. 1619–1627, 2012.

[15]   K. Fukui and O. Yamaguchi, "Face recognition using multi-viewpoint patterns for robot vision," in *Proc. 11th Int. Symp. Robotics Res.*, 2003, pp. 192–201.

[16]   J. R. Beveridge, B. A. Draper, J.-M. Chang, M. Kirby, H. Kley, and C. Peterson, "Principal angles separate subject illumination spaces in ydb and cmu-pie," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 31, no. 2, pp. 351–363, Feb. 2009.

[17]   H. Hotelling, "Relation between two sets of variables," *Biometrica*, vol. 28, pp. 322–377, 1936.

[18]   S. Afriat, "Orthogonal and oblique projectors and the characteristics of pairs of vector spaces," in *Proc. Cambridge Philos. Soc.*, 1957, vol. 53, pp. 800–816.

[19]   M. Turk and A. Pentland, "Face recognition using eigenfaces," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, 1993, pp. 453–458.

[20]   P. N. Belhumeur, J. P. Hespanha, and D. J. Kriegman, "Eigenfaces vs. fisherfaces: Recognition using class specific linear projection," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 19, no. 7, pp. 711–720, Jul. 1997.

[21]   K. Fukunaga and W. Koontz, "Application of the Karhunen-Loeve expansion to feature selection and ordering," *IEEE Trans. Comput.*, vol. 19, no. 4, pp. 311–318, Apr. 1970.

[22]   K. Fukunaga, *Introduction to Statistical Pattern Recognition*. New York, NY, USA: Academic, 1972.

[23]   B. Schölkopf, A. Smola, and K.-R. Müller, "Nonlinear principal component analysis as a kernel eigenvalue problem," *Neural Comput.*, vol. 10, pp. 1299–1319, 1998.

[24]   E. Maeda and H. Murase, "Multi-category classification by kernel based nonlinear subspace method," in *Proc. IEEE Int. Conf. Acoust., Speech, Signal Process.*, 1999, vol. 2, pp. 1025–1028.

[25]   K. Tsuda, "Subspace classifier in the Hilbert space," *Pattern Recog. Lett.*, vol. 20, pp. 513–519, 1999.

[26]   L. Wolf and A. Shashua, "Learning over sets using kernel principal angles," *J. Mach. Learn. Res.*, vol. 4, pp. 913–931, 2003.

[27]   K. Fukui, B. Stenger, and O. Yamaguchi, "A framework for 3d object recognition using the kernel constrained mutual subspace method," in *Proc. Asian Conf. Comput. Vis.*, 2006, pp. 315–324.

[28]   J. Kittler, "The subspace approach to pattern recognition," *Progress Cybern. Syst. Res.*, pp. 92–97, 1978.

[29]   T. Kawahara, M. Nishiyama, T. Kozakaya, and O. Yamaguchi, "Face recognition based on whitening transformation of distribution of subspaces," in *Proc. ACCV Workshops, Subspace*, 2007, pp. 97–103.

[30]   S. Zhang and T. Sim, "Discriminant subspace analysis: A Fukunaga-Koontz approach," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 29, no. 10, pp. 1732–1745, Oct. 2007.

[31]   K. Fukui and O. Yamaguchi, "The kernel orthogonal mutual subspace method and its application to 3d object recognition," in *Proc. Asian Conf. Comput. Vis.*, 2007, pp. 467–476.

[32]   J. Hamm and D. D. Lee, "Grassmann discriminant analysis: A unifying view on subspace-based learning," in *Proc. Int. Conf. Mach. Learn.*, 2008, pp. 376–383.

[33]   P. Turaga, A. Veeraraghavan, A. Srivastava, and R. Chellappa, "Statistical computations on Grassmann and Stiefel manifolds for image and video-based recognition," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 33, no. 11, pp. 2273–2286, Nov. 2011.

[34]   Y. Chikuse, *Statistics on Special Manifolds*. New York, NY, USA: Springer, vol. 174, 2013.

[35]   H. Yanai, "Some generalized forms a least squares g-inverse, minimum norm g-inverse, and Moore–Penrose inverse matrices," *Comput. Statist. Data Anal.*, vol. 10, pp. 251–260, 1990.

[36]   R. Gross, I. Matthews, J. F. Cohn, T. Kanade, and S. Baker, "Multipie," *Image Vis. Comput.*, vol. 28, no. 5, pp. 807–813, 2010.

[37]   B. Schölkopf, A. J. Smola, P. Knirsch, and C. Burges, "Fast approximation of support vector kernel expansions, and an interpretation of clustering as approximation in feature spaces," in *Proc. Mustererkennung 20. DAGM-Symp.*, 1998, pp. 125–132.

[38]   H. F. Chen, P. N. Belhumeur, and D. W. Jacobs, "In search of illumination invariants," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, 2000, pp. 1254–1261.

[39]   T. K. Kim, J. Kittler, and R. Cipolla, "On-line learning of mutually orthogonal subspaces for face recognition by image sets," *IEEE Trans. Image Process.*, vol. 19, no. 4, pp. 1067–1074, Apr. 2010.

[40]   J. Wright, A. Y. Yang, A. Ganesh, S. S. Sastry, and Y. Ma, "Robust face recognition via sparse representation," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 31, no. 2, pp. 210–227, Feb. 2009.

Kazuhiro Fukui received the BE and ME degrees in mechanical engineering from Kyushu University, in 1986 and 1988, respectively, and the PhD degree from the Tokyo Institute of Technology in 2003. In 1988, he joined Toshiba Corporate R&D Center and served as a senior research scientist at Multimedia Laboratory, in 2002. He is currently a professor in the Department of Computer Science, Graduate School of Systems and Information Engineering at the University of Tsukuba. His interests include the theory of computer vision, pattern recognition, and applications of these theories. He has been serving as a program committee member at many pattern recognition and computer vision conferences, including as an area chair of ICPR'12 and ICPR'14. He is a member of the IEEE and SIAM.

Atsuto Maki received the PhD degree in computing science from the Royal Institute of Technology (KTH), Sweden, in 1996. Prior to that, he studied electrical engineering at Kyoto University and the University of Tokyo. He is currently an associate professor at the School of Computer Science and Communication, KTH. He has previously been a senior researcher at Toshiba Research Cambridge, U.K., an associate professor at the Graduate School of Informatics, Kyoto University, and with Toshiba Corporate R&D Center. His interests cover a broad range of topics in machine learning and computer vision, including motion and object recognition, clustering, representation learning, and stereopsis. He has been serving as a program committee member at major computer vision conferences, including as an area chair of ICCV'09. He is a member of the IEEE.