

# Generalized Mutual Subspace Based Methods for Image Set Classification

Takumi Kobayashi

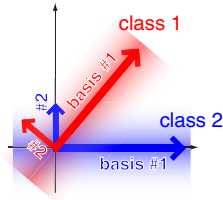
National Institute of Advanced Industrial Science and Technology, Japan  
takumi.kobayashi@aist.go.jp

**Abstract.** The subspace-based methods are effectively applied to classify *sets* of feature vectors by modeling them as subspaces. It is, however, difficult to appropriately determine the subspace dimensionality in advance for better performance. For alleviating such issue, we present a generalized mutual subspace method by introducing *soft weighting* across the basis vectors of the subspace. The bases are effectively combined via the soft weights to measure the subspace similarities (angles) without definitely setting the subspace dimensionality. By using the soft weighting, we consequently propose a novel mutual subspace-based method to construct the discriminative space which renders more discriminative subspace similarities. In the experiments on 3D object recognition using image sets, the proposed methods exhibit stably favorable performances compared to the other subspace-based methods.

## 1 Introduction

In recent visual recognition tasks, *sets* of images (feature vectors) are effectively employed; *e.g.*, image frames in a video sequence for face recognition [1–4] and multiple still images captured from various angles for 3D object recognition [5, 6]. Those image sets capture various appearance changes of objects, providing more discriminative clues for classification than a single-shot image alone. Thus, the recognition systems utilizing those sets exhibit superior performances [1, 7].

The subspace models are successfully applied to classify the sets of vectors [5–13]. The vector set is represented by its underlying subspace spanned by a small number of the principal basis vectors. The angle between a vector and a subspace [14] is mathematically extended to the canonical angle between subspaces [7, 15, 16] which is a fundamental measurement for classifying the sets. The canonical angles measured in the original vector space, however, are not necessarily favorable from the viewpoint of discriminating classes, and for improving classification performances, apart from kernelization [6, 13], it is important to discriminatively measure those canonical angles. For that purpose, much research effort has been made for constructing the discriminative space [5, 8–11]. Those methods render the space such that the embedded subspaces are discriminatively separated in terms of the canonical angles, as in Fisher discriminant analysis [17] which makes the pair-wise vector distance more discriminative.



**Fig. 1.** The arrows indicate the directions of the bases with the variances (arrow length) in two classes. The one-dimensional subspaces (longer arrows) discriminate the two class, but they coincide when employing two-dimensional subspace. The proposed method can properly distinguish them with soft weighting based on the variances.

The subspace-based methods, however, involve the crucial issue regarding the dimensionality of the subspace. Users are required in advance to appropriately determine the subspace dimensionality for better performance. Each of the basis vectors in the set, given by an eigen-decomposition (PCA), is subject to the decision whether it is employed to support the subspace or not, based on the eigenvalues (variances). Such *binary* decision may significantly affect the performance as shown in Fig. 1, and the proper number of the bases, *i.e.*, the subspace dimensionality, is not known a priori depending on the overlap among the classes. The importance of the bases should be defined smoothly according to the eigenvalues (variances) rather than in such a binary form.

We propose generalized mutual subspace based methods for alleviating the issue regarding the subspace dimensionality. The proposed methods introduce *soft weighting* on the basis vectors composing the subspace, without definitely picking up a small number of the principal bases. We generalize the mutual subspace method (MSM) [7] by using the soft weighting to effectively combine the bases for computing subspace angles. In addition, we reformulate the constrained mutual subspace method (CMSM) [8, 9] constructing the discriminative space and give it theoretical justification. The reformulation also enables us to generalize the method by incorporating the soft weighting. The proposed methods effectively exploit all the bases, some of which are even less contributive to support the subspace, via the soft weighting. In summary, our contributions are 1) to introduce the soft weighting across the bases for effectively computing the subspace angles in MSM, 2) to reformulate CMSM with theoretical justification and 3) to propose a novel method for constructing the discriminative space based on the reformulation with the soft weighting.

This paper is organized as follows: in the next section, we briefly review the methods of MSM and CMSM and describe the novel reformulation of CMSM. In Sec.3, the details of the proposed methods are described and Sec.4 shows the experimental results on 3D object recognition. We conclude this paper in Sec.5.

## 2 Subspace based methods

The subspace model is effectively applied to classify not only a feature vector [14] but also a set of feature vectors [7] that is our main concern in this paper. The

set of feature vectors is represented by the subspace and thereby samples are those subspaces, not vectors.

## 2.1 Mutual subspace method (MSM)

Let the  $i$ -th set of feature vectors be denoted by  $\mathbf{X}_i = [\mathbf{x}_{i1}, \dots, \mathbf{x}_{in_i}] \in \mathbb{R}^{d \times n_i}$  where  $\mathbf{x}_{ij} \in \mathbb{R}^d$  is the  $j$ -th  $d$ -dimensional feature vector in the  $i$ -th set. The set of vectors is modeled by the subspace spanned by the principal basis vectors: we first apply an eigen-decomposition to  $\mathbf{X}_i \mathbf{X}_i^\top = \mathbf{V}_i \mathbf{\Lambda}_i \mathbf{V}_i^\top$  where  $\mathbf{V}_i, \mathbf{\Lambda}_i = \text{diag}(\{\lambda_{il}\}_{l=1}^d) \in \mathbb{R}^{d \times d}$  are the eigenvectors and the eigenvalue diagonal matrix, respectively, and then exploit the eigenvectors of the  $m_i$  larger eigenvalues as the orthonormal principal vectors  $\hat{\mathbf{V}}_i \in \mathbb{R}^{d \times m_i}$ .

To classify the set of vectors, *i.e.*, the subspace, the similarity measure is defined by the canonical angles [15, 16] between the subspaces, called mutual subspace method (MSM) [7]:

$$\cos \theta_{ij}^{(k)} = \max_{\mathbf{q}_{ik} \in \mathbb{R}^{m_i}, \mathbf{q}_{jk} \in \mathbb{R}^{m_j}} \mathbf{q}_{ik}^\top \hat{\mathbf{V}}_i^\top \hat{\mathbf{V}}_j \mathbf{q}_{jk}, \quad k = 1, \dots, \min[m_i, m_j] \quad (1)$$

*s.t.*,  $\|\mathbf{q}_{ik}\| = \|\mathbf{q}_{jk}\| = 1, \mathbf{q}_{ik}^\top \mathbf{q}_{ik'} = \mathbf{q}_{jk}^\top \mathbf{q}_{jk'} = 0, k \neq k'$ .

This is computed by applying singular value decomposition (SVD) to  $\hat{\mathbf{V}}_i^\top \hat{\mathbf{V}}_j$ :

$$\hat{\mathbf{V}}_i^\top \hat{\mathbf{V}}_j = \mathbf{Q}_i \mathbf{\Theta}_{ij} \mathbf{Q}_j^\top, \quad (2)$$

where  $\mathbf{Q}_i, \mathbf{Q}_j$  are the singular vectors composed of  $\mathbf{q}_{ik}, \mathbf{q}_{jk}$  and  $\mathbf{\Theta}_{ij}$  is the singular-value diagonal matrix whose diagonal elements are the singular values  $\{\cos \theta_{ij}^{(k)}\}_k$ . The  $k$ -th canonical vectors  $\hat{\mathbf{V}}_i \mathbf{q}_{ik}$  and  $\hat{\mathbf{V}}_j \mathbf{q}_{jk}$  provide the  $k$ -th canonical angle  $\cos \theta_{ij}^{(k)}$ . In this paper, we define the similarity measure by  $s_{ij} = \text{tr}(\mathbf{\Theta}_{ij}^2)$ . The measure  $s_{ij}$  becomes high for the similar subspaces and low for the distinct ones, and they are fed to  $k$ -NN classifications.

It should be noted that in MSM the subspace dimensionality  $m_i$  is crucial for better classification performances, since the larger  $m_i$  causes the measure  $s_{ij}$  to be uniformly high for any pairs of subspaces;  $s_{ij} = d, \forall i, j$  in case of  $m_i = d$ . There is no theoretical way to appropriately determine the dimensionality, and thus users are required in advance to carefully tune the dimensionality, which is an exhaustive procedure from the practical viewpoint.

## 2.2 Constrained mutual subspace method (CMSM)

In the above section, we have described how MSM works on classifying sets of feature vectors in the original vector space. The original space, however, is not necessarily favorable in terms of discriminating the subspaces, and it is desirable to embed those subspaces into the more discriminative space as in Fisher discriminant analysis (FDA) [17]. For that purpose, some methods have been proposed for constructing the discriminative space to make the subspace angles measured in that space more discriminative [5, 8–11]. In this paper, we focus on the constrained mutual subspace method (CMSM) [8, 9] due to its simple formulation and promising performances.

Each class is represented by the subspace  $\hat{\mathbf{V}}^{[c]} \in \mathfrak{R}^{d \times m_c}$ ,  $c = 1, \dots, C$ , and let the collection of those subspace bases be denoted by  $\bar{\mathbf{V}} = [\hat{\mathbf{V}}^{[1]}, \dots, \hat{\mathbf{V}}^{[C]}]$ . CMSM produces the projection vectors  $\mathbf{D}$  by the eigenvectors of the smaller eigenvalues in the following eigenvalue problem:

$$\bar{\mathbf{V}} \bar{\mathbf{V}}^\top \mathbf{D} = \mathbf{D} \boldsymbol{\Sigma}, \text{ s.t.}, \mathbf{D}^\top \mathbf{D} = \mathbf{I}, \quad (3)$$

where  $\boldsymbol{\Sigma} = \text{diag}(\{\sigma_l\}_l)$  is the eigenvalue diagonal matrix. In [8], it is shown that those projection vectors of the smaller eigenvalues are based on the differential vectors between the canonical vectors, which discriminatively separate the subspaces.

In a more practical case that multiple subspaces  $\{\hat{\mathbf{V}}_i^{[c]}\}_{i=1}^{n_c}$  are given in each class, a heuristic approach is presented in [9]: CMSM is repeatedly applied to *within-class* subspaces and *between-class* ones as follows.

1. **Within-class:** In the  $c$ -th class, the representative subspace  $\mathbf{P}^{[c]}$  is obtained by applying CMSM to  $\bar{\mathbf{V}}^{[c]} = [\hat{\mathbf{V}}_1^{[c]}, \dots, \hat{\mathbf{V}}_{n_c}^{[c]}]$  as  $\bar{\mathbf{V}}^{[c]} \bar{\mathbf{V}}^{[c]\top} = \mathbf{P}^{[c]} \boldsymbol{\Sigma}^{[c]} \mathbf{P}^{[c]\top}$ .
2. **Between-class:** CMSM is applied to  $\bar{\mathbf{V}} = [\mathbf{P}^{[1]}, \dots, \mathbf{P}^{[C]}]$  in (3) to get the projection vectors  $\mathbf{D}$ .

### 2.3 Reformulation of CMSM

In this paper, we give justification to CMSM in a different way from [8, 9] by reformulating (3); the proposed method is founded on the reformulation (see Sec.3). We consider the following dual eigenvalue problem of (3):

$$\bar{\mathbf{V}}^\top \bar{\mathbf{V}} \mathbf{Q} = \mathbf{Q} \boldsymbol{\Sigma}, \text{ s.t.}, \mathbf{Q}^\top \mathbf{Q} = \mathbf{I}, \quad (4)$$

and the projection matrix  $\mathbf{D}$  is retrieved by

$$\mathbf{D} = \bar{\mathbf{V}} \mathbf{Q} \boldsymbol{\Sigma}^{-\frac{1}{2}}, \quad (5)$$

since  $\bar{\mathbf{V}} = \mathbf{D} \boldsymbol{\Sigma}^{\frac{1}{2}} \mathbf{Q}^\top$  (SVD). The novel formulation (4) provides the coefficients  $\mathbf{Q}$  on the basis vectors, like coefficients on support vectors in SVM [18], to construct discriminative space. We can give two interpretations (justifications) to (4) as follows.

- 1) **Differential canonical vectors.** The reformulation (4) simply shows that the projection vectors are based on the differential vectors between canonical vectors. Suppose the two class subspaces  $\hat{\mathbf{V}}_1, \hat{\mathbf{V}}_2$  and  $\bar{\mathbf{V}} = [\hat{\mathbf{V}}_1, \hat{\mathbf{V}}_2]$ . The dual eigenvalue problem (4) is described by

$$\begin{aligned} \bar{\mathbf{V}}^\top \bar{\mathbf{V}} \frac{1}{\sqrt{2}} \begin{bmatrix} \mathbf{Q}_1 & \mathbf{Q}_1 \\ \mathbf{Q}_2 & -\mathbf{Q}_2 \end{bmatrix} &= \begin{bmatrix} \mathbf{I} & \hat{\mathbf{V}}_1^\top \hat{\mathbf{V}}_2 \\ \hat{\mathbf{V}}_2^\top \hat{\mathbf{V}}_1 & \mathbf{I} \end{bmatrix} \frac{1}{\sqrt{2}} \begin{bmatrix} \mathbf{Q}_1 & \mathbf{Q}_1 \\ \mathbf{Q}_2 & -\mathbf{Q}_2 \end{bmatrix} \\ &= \frac{1}{\sqrt{2}} \begin{bmatrix} \mathbf{Q}_1 + \mathbf{Q}_1 \boldsymbol{\Theta}_{12} & \mathbf{Q}_1 - \mathbf{Q}_1 \boldsymbol{\Theta}_{12} \\ \mathbf{Q}_2 \boldsymbol{\Theta}_{12} + \mathbf{Q}_2 & \mathbf{Q}_2 \boldsymbol{\Theta}_{12} - \mathbf{Q}_2 \end{bmatrix} = \frac{1}{\sqrt{2}} \begin{bmatrix} \mathbf{Q}_1 & \mathbf{Q}_1 \\ \mathbf{Q}_2 & -\mathbf{Q}_2 \end{bmatrix} \begin{bmatrix} \mathbf{I} + \boldsymbol{\Theta}_{12} & \mathbf{0} \\ \mathbf{0} & \mathbf{I} - \boldsymbol{\Theta}_{12} \end{bmatrix}. \end{aligned}$$

where  $\mathbf{Q}_1, \mathbf{Q}_2, \boldsymbol{\Theta}_{12}$  are the singular vectors and the singular values in (2). Thus, the projection vectors of the smaller eigenvalues are  $\mathbf{D} = \frac{1}{\sqrt{2}} (\hat{\mathbf{V}}_1 \mathbf{Q}_1 - \hat{\mathbf{V}}_2 \mathbf{Q}_2) (\mathbf{I} - \boldsymbol{\Theta}_{12})^{-\frac{1}{2}}$ , which are (normalized) differential vectors between the canonical vectors,  $\hat{\mathbf{V}}_1 \mathbf{Q}_1$  and  $\hat{\mathbf{V}}_2 \mathbf{Q}_2$ , of those two classes. This shows that

the eigenvalues less than 1 exhibit the discriminative power, rendering the differential canonical vectors for the projections.

- 2) **Minimization of pair-wise canonical angles.** The eigenvalue problem (4) is also derived from the following optimization:

$$\min_{\|\mathbf{q} \triangleq [\mathbf{q}^{[1]\top}, \dots, \mathbf{q}^{[C]\top}]^\top\| = 1} \sum_{c, c' | c \neq c'}^C \mathbf{q}^{[c]\top} \hat{\mathbf{V}}^{[c]\top} \hat{\mathbf{V}}^{[c']} \mathbf{q}^{[c']} + \mathbf{q}^\top \mathbf{q} = \mathbf{q}^\top \bar{\mathbf{V}}^\top \bar{\mathbf{V}} \mathbf{q}, \quad (6)$$

where we use  $\hat{\mathbf{V}}^{[c]\top} \hat{\mathbf{V}}^{[c]} = \mathbf{I}$ ,  $\forall c$ . The second term  $\mathbf{q}^\top \mathbf{q} (= 1)$  is the same as the constraint, which thus makes no effect on the optimization. Roughly speaking, the first term  $J \triangleq \sum_{c \neq c'} \mathbf{q}^{[c]\top} \hat{\mathbf{V}}^{[c]\top} \hat{\mathbf{V}}^{[c']} \mathbf{q}^{[c']}$  amounts to sum of canonical angles between different classes  $c$  and  $c'$  as in (1). The eigenvalue less than 1 indicates  $J < 0$  in which the discriminative coefficients  $\mathbf{q}$  are extracted.

The another merit of the reformulation (4) is that we can give theoretical justification to the heuristic procedure [9] described in Sec.2.2 for dealing with multiple subspaces in each class. Based on the above-mentioned second interpretation, we consider the optimization problem to minimize the canonical angles between classes while maximizing those within classes, which is formulated as in FDA [17] by

$$J = \min_{\mathbf{q}} \frac{\sum_{c, c' | c \neq c'}^C \sum_i^{n_c} \sum_j^{n_{c'}} \mathbf{q}_i^{[c]\top} \hat{\mathbf{V}}_i^{[c]\top} \hat{\mathbf{V}}_j^{[c']} \mathbf{q}_j^{[c']}}{\sum_c^C \sum_{i, j | i \neq j}^{n_c} \mathbf{q}_i^{[c]\top} \hat{\mathbf{V}}_i^{[c]\top} \hat{\mathbf{V}}_j^{[c]} \mathbf{q}_j^{[c]} + \mathbf{q}^\top \mathbf{q}} = \min_{\mathbf{q}} \frac{\mathbf{q}^\top \mathbf{R}_B \mathbf{q}}{\mathbf{q}^\top \mathbf{R}_W \mathbf{q}} \quad (7)$$

$$\Rightarrow \mathbf{R}_B \mathbf{Q} = \mathbf{R}_W \mathbf{Q} \tilde{\Sigma} \Leftrightarrow \mathbf{R}_T \mathbf{Q} = \mathbf{R}_W \mathbf{Q} \Sigma, \quad s.t., \mathbf{Q}^\top \mathbf{Q} = \mathbf{I}, \quad (8)$$

where  $\Sigma = \tilde{\Sigma} + \mathbf{I}$  and

$$\mathbf{R}_W = \begin{bmatrix} \bar{\mathbf{V}}^{[1]\top} \bar{\mathbf{V}}^{[1]} & & & \\ & \ddots & & \\ & & \bar{\mathbf{V}}^{[C]\top} \bar{\mathbf{V}}^{[C]} & \\ & & & \end{bmatrix} = \text{blkdiag}(\{\bar{\mathbf{V}}^{[c]\top} \bar{\mathbf{V}}^{[c]}\}_{c=1}^C) \succcurlyeq 0,$$

$$\mathbf{R}_T = [\bar{\mathbf{V}}^{[1]}, \dots, \bar{\mathbf{V}}^{[C]}]^\top [\bar{\mathbf{V}}^{[1]}, \dots, \bar{\mathbf{V}}^{[C]}] \succcurlyeq 0, \quad \mathbf{R}_B = \mathbf{R}_T - \mathbf{R}_W,$$

where  $\text{blkdiag}$  constructs a block-diagonal matrix. The projection matrix is given in a manner similar to (5) by

$$\mathbf{D} = [\bar{\mathbf{V}}^{[1]}, \dots, \bar{\mathbf{V}}^{[C]}] \mathbf{Q} \Sigma^{-\frac{1}{2}}. \quad (9)$$

As in (6),  $\mathbf{q}^\top \mathbf{R}_B \mathbf{q}$  and  $\mathbf{q}^\top \mathbf{R}_W \mathbf{q}$  measure *between-class* and *within-class* canonical angles, respectively. This proposed formulation has the following properties with respect to the projection vectors  $\mathbf{D}$  and the eigenvalues  $\Sigma = \text{diag}\{\sigma_l\}_l$ .

**Theorem 1.** *The projection vectors (9) produced via (8) are the same as the projections by the heuristic procedure [9] described in Sec.2.2.*

*Proof.* We apply SVD to  $\bar{\mathbf{V}}^{[c]} = \mathbf{P}^{[c]} \Sigma^{[c]} \mathbf{Q}^{[c]\top}$ , and define a block-diagonal matrix  $\mathbf{R}_W^{-\frac{1}{2}} = \text{blkdiag}(\{\mathbf{Q}^{[c]} \Sigma^{[c]-1}\}_{c=1}^C)$  such that  $\mathbf{R}_W^{-\frac{1}{2}\top} \mathbf{R}_W \mathbf{R}_W^{-\frac{1}{2}} = \mathbf{I}$ . By using these, (8) is transformed into

$$\mathbf{R}_W^{-\frac{1}{2}\top} \mathbf{R}_T \mathbf{R}_W^{-\frac{1}{2}} \mathbf{Q}' = \mathbf{Q}' \Sigma, \quad s.t., \mathbf{Q}'^\top \mathbf{Q}' = \mathbf{I}, \quad \text{where } \mathbf{Q} = \mathbf{R}_W^{-\frac{1}{2}} \mathbf{Q}'. \quad (10)$$

The left-hand-side matrix results in

$$\begin{aligned} \mathbf{R}_W^{-\frac{1}{2}\top} \mathbf{R}_T \mathbf{R}_W^{-\frac{1}{2}} &= \mathbf{R}_W^{-\frac{1}{2}\top} [\bar{\mathbf{V}}^{[1]}, \dots, \bar{\mathbf{V}}^{[C]}\top [\bar{\mathbf{V}}^{[1]}, \dots, \bar{\mathbf{V}}^{[C]}] \mathbf{R}_W^{-\frac{1}{2}} \\ &= [\mathbf{P}^{[1]}, \dots, \mathbf{P}^{[C]}\top [\mathbf{P}^{[1]}, \dots, \mathbf{P}^{[C]}]. \end{aligned} \quad (11)$$

This is the same procedure as within-class CMSM to obtain the class representative subspaces  $\mathbf{P}^{[c]}$  in [9].  $\bar{\mathbf{V}} = [\bar{\mathbf{V}}^{[1]}, \dots, \bar{\mathbf{V}}^{[C]}]$  is decomposed via SVD by

$$\bar{\mathbf{V}} = \mathbf{D} \boldsymbol{\Sigma}^{\frac{1}{2}} \mathbf{Q}'\top,$$

where  $\mathbf{D}$  corresponds to the projection vectors heuristically given in [9] and  $\mathbf{Q}'$  is the eigenvectors in (10). The projection vectors (9) are finally described by

$$\begin{aligned} \mathbf{D}^{dual} &= [\bar{\mathbf{V}}^{[1]}, \dots, \bar{\mathbf{V}}^{[C]}] \mathbf{Q} \boldsymbol{\Sigma}^{-\frac{1}{2}} = [\bar{\mathbf{V}}^{[1]}, \dots, \bar{\mathbf{V}}^{[C]}] \mathbf{R}_W^{-\frac{1}{2}} \mathbf{Q}' \boldsymbol{\Sigma}^{-\frac{1}{2}} \\ &= [\mathbf{P}^{[1]}, \dots, \mathbf{P}^{[C]}] \mathbf{Q}' \boldsymbol{\Sigma}^{-\frac{1}{2}} = \bar{\mathbf{V}} \mathbf{Q}' \boldsymbol{\Sigma}^{-\frac{1}{2}} = \mathbf{D}. \quad \square \end{aligned}$$

**Theorem 2.** *The eigenvalues in (8) are bounded in  $0 \leq \sigma_l \leq C$ ,  $\forall l$ .*

*Proof.*  $\mathbf{R}_T \succcurlyeq 0$  and  $\mathbf{R}_W \succcurlyeq 0$  lead to  $\sigma_l \geq 0$ ,  $\forall l$ . By (10) and (11), the maximum eigenvalue is obtained by

$$\max_l \sigma_l = \max_{\|\mathbf{q}\|=1} \mathbf{q}'\top [\mathbf{P}^{[1]}, \dots, \mathbf{P}^{[C]}\top [\mathbf{P}^{[1]}, \dots, \mathbf{P}^{[C]}] \mathbf{q}, \quad (12)$$

where  $\mathbf{P}^{[c]}$  is the orthonormal vectors since  $\bar{\mathbf{V}}^{[c]} = \mathbf{P}^{[c]} \boldsymbol{\Lambda}^{[c]} \mathbf{Q}^{[c]\top}$  (SVD). Here, we consider  $\mathbf{P}^{[c]} \mathbf{q}^{[c]} = \bar{q}^{[c]} \bar{\mathbf{p}}^{[c]}$  where  $\|\bar{\mathbf{p}}^{[c]}\| = 1$  and  $\bar{q}^{[c]} = \sqrt{\mathbf{q}^{[c]\top} \mathbf{q}^{[c]}}$ , and  $[\mathbf{P}^{[1]}, \dots, \mathbf{P}^{[C]}] \mathbf{q} = [\bar{\mathbf{p}}^{[1]}, \dots, \bar{\mathbf{p}}^{[C]}] \bar{\mathbf{q}}$  where  $\bar{\mathbf{q}} = [\bar{q}^{[1]}, \dots, \bar{q}^{[C]}\top \in \Re^C$ . Thus, the right-hand side in (12) is bounded in

$$\bar{\mathbf{q}}\top [\bar{\mathbf{p}}^{[1]}, \dots, \bar{\mathbf{p}}^{[C]}\top [\bar{\mathbf{p}}^{[1]}, \dots, \bar{\mathbf{p}}^{[C]}] \bar{\mathbf{q}} \leq \bar{\mathbf{q}}\top \mathbf{1} \mathbf{1}\top \bar{\mathbf{q}} \leq C, \quad s.t. \|\bar{\mathbf{q}}\| = 1,$$

where we use  $\|\bar{\mathbf{q}}\| = \|\mathbf{q}\| = 1$  and  $\bar{\mathbf{p}}^{[c]\top} \bar{\mathbf{p}}^{[c]} \leq 1$ ,  $\forall c, c'$ .  $\square$

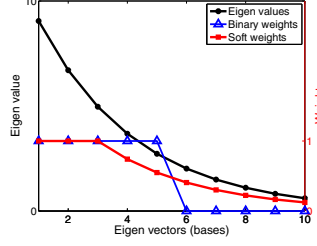
### 3 Proposed methods

As mentioned in Sec.2.1, the crucial issue in the subspace-based methods is how to determine the dimensionality of the subspace. For alleviating it, the main idea in this paper is to introduce *soft weighting* across the basis vectors of the subspace instead of definitely selecting a small number of the principal bases.

#### 3.1 Generalized mutual subspace method (gMSM)

The determination of the subspace dimensionality is regarded as picking up the principal basis vectors, which further corresponds to designing the binary weights on all the bases as shown in Fig. 2; 1/0 indicates whether the basis is picked up or not. We relax the binary (hard) weighting to *soft weighting*. Let  $\boldsymbol{\omega} \in [0, 1]^d$  be the soft weights on the bases and  $\boldsymbol{\Omega} = \text{diag}(\boldsymbol{\omega}) \in \Re^{d \times d}$ . The soft weight is illustrated in Fig. 2 compared to the binary one. By incorporating the weights into MSM in (1), the generalized MSM (gMSM) is defined by

$$\max_{\mathbf{q}_i\top \boldsymbol{\Omega}_i^{-2} \mathbf{q}_i = 1, \mathbf{q}_j\top \boldsymbol{\Omega}_j^{-2} \mathbf{q}_j = 1} \mathbf{q}_i\top \mathbf{V}_i\top \mathbf{V}_j \mathbf{q}_j \Leftrightarrow \max_{\mathbf{q}'_i\top \mathbf{q}'_i = 1, \mathbf{q}'_j\top \mathbf{q}'_j = 1} \mathbf{q}'_i\top \boldsymbol{\Omega}_i \mathbf{V}_i\top \mathbf{V}_j \boldsymbol{\Omega}_j \mathbf{q}'_j, \quad (13)$$



**Fig. 2.** Weights on basis vectors. The binary weight indicates five dimensions of the subspace and the soft weight is created by  $w_3$  in (15).

where  $\mathbf{q}_i = \mathbf{\Omega}_i \mathbf{q}'_i$  and  $\mathbf{V}_i$  indicates the orthonormal bases given by an eigen-decomposition of  $\mathbf{X}_i \mathbf{X}_i^\top = \mathbf{V}_i \mathbf{\Lambda}_i \mathbf{V}_i^\top$ . As in (2), the generalized canonical angles are computed by

$$\mathbf{\Omega}_i \mathbf{V}_i^\top \mathbf{V}_j \mathbf{\Omega}_j = \mathbf{Q}'_i \mathbf{\Theta}_{ij} \mathbf{Q}'_j{}^\top, \quad s_{ij} = \text{tr}(\mathbf{\Theta}_{ij}^2), \quad (14)$$

and the  $k$ -th canonical vectors are described by  $\mathbf{V}_i \mathbf{q}_i = \mathbf{V}_i \mathbf{\Omega}_i \mathbf{q}'_i$  and  $\mathbf{V}_j \mathbf{\Omega}_j \mathbf{q}'_j$ . In this formulation for measuring the generalized canonical angle, the soft weights on the bases work in the constraint  $\mathbf{q}^\top \mathbf{\Omega}^{-2} \mathbf{q} = 1$ ; the smaller weight  $\omega_l$  decreases the coefficient  $q_l$ , and especially  $\omega_l \rightarrow 0$  enforces  $q_l \rightarrow 0$ . The gMSM in (13) obviously reduces to the ordinary MSM in (1) by  $\boldsymbol{\omega} \in \{0, 1\}^d$  (binary weights) since  $[\hat{\mathbf{V}}_i, \mathbf{0}] = \mathbf{V}_i \mathbf{\Omega}_i$ . While in MSM the classification performance is, like quantization errors, sensitive to the dimensionality indicated by the binary weights, the soft weights in gMSM effectively combine the bases to compute the subspace angles.

The only issue is how to design the soft weights. Since the dimensionality is usually determined based on the variances corresponding to the eigenvalues in  $\mathbf{X} \mathbf{X}^\top = \mathbf{V} \mathbf{\Lambda} \mathbf{V}^\top$ , the soft weights are also set according to the eigenvalues by

$$\boldsymbol{\omega} = \mathbf{w}_m(\boldsymbol{\lambda}), \quad \text{where } \mathbf{\Lambda} = \text{diag}(\boldsymbol{\lambda}) \text{ and } w_m(\lambda) = \min \left[ \frac{\lambda}{\lambda_m}, 1 \right], \quad (15)$$

where  $\lambda_m$  is the  $m$ -th eigenvalue in  $\boldsymbol{\lambda} \in \Re^d$  in descending order and  $\min$  operates on each component of the vector in comparison to 1, enforcing  $\omega \leq 1$ . This weighting evaluates the importance of the basis vector by the variance relative to  $\lambda_m$ ; the basis of the larger variance is more important. In most cases, the first principal basis vector indicate the direction of the subspace, which results in significantly large  $\lambda_1$ , and the second or later ones capture the spread of the distribution around that direction. Thus, we suggest to use  $m > 1$ ; the soft weights with  $m=3$  are illustrated in Fig. 2.

### 3.2 Generalized constrained mutual subspace method (gCMSM)

Along with gMSM, CMSM is also generalized by incorporating the soft weights into the optimization problem (6) of the reformulated CMSM:

$$\min_{\mathbf{q}^\top \bar{\mathbf{\Omega}}^{-2} \mathbf{q} = 1} \sum_{c, c' | c \neq c'}^C \mathbf{q}^{[c]\top} \mathbf{V}^{[c]\top} \mathbf{V}^{[c']} \mathbf{q}^{[c']} + \mathbf{q}^\top \bar{\mathbf{\Omega}}^{-2} \mathbf{q} \quad (16)$$

$$\Leftrightarrow \min_{\mathbf{q}^\top \bar{\Omega}^{-2} \mathbf{q} = 1} \mathbf{q}^\top (\mathbf{R}_T - \mathbf{I} + \bar{\Omega}^{-2}) \mathbf{q} \Leftrightarrow \min_{\mathbf{q}'^\top \mathbf{q}' = 1} \mathbf{q}'^\top (\bar{\Omega} \mathbf{R}_T \bar{\Omega} - \bar{\Omega}^2 + \mathbf{I}) \mathbf{q}', \quad (17)$$

where  $\mathbf{q} \triangleq [\mathbf{q}^{[1]\top}, \dots, \mathbf{q}^{[C]\top}]^\top = \bar{\Omega} \mathbf{q}'$ ,  $\bar{\Omega} = \text{blkdiag}(\{\Omega^{[c]}\}_{c=1}^C)$  and  $\mathbf{R}_T = \bar{\mathbf{V}}^\top \bar{\mathbf{V}}$ ,  $\bar{\mathbf{V}} = [\mathbf{V}^{[1]}, \dots, \mathbf{V}^{[C]}]$ . This results in the following eigenvalue problem:

$$(\bar{\Omega} \mathbf{R}_T \bar{\Omega} - \bar{\Omega}^2 + \mathbf{I}) \mathbf{Q}' = \mathbf{Q}' \Sigma, \quad \text{s.t.}, \mathbf{Q}'^\top \mathbf{Q}' = \mathbf{I}, \quad (18)$$

and by using the eigenvectors of the smaller eigenvalues, the projection vectors are obtained by

$$\mathbf{D} = \bar{\mathbf{V}} \mathbf{Q} \Sigma^{-\frac{1}{2}} = \bar{\mathbf{V}} \bar{\Omega} \mathbf{Q}' \Sigma^{-\frac{1}{2}}. \quad (19)$$

As in the ordinary CMSM, the gCMSM also implies the differential vectors between the canonical vectors as follows. Suppose two class bases  $\mathbf{V}_1, \mathbf{V}_2$  and  $\bar{\mathbf{V}} = [\mathbf{V}_1, \mathbf{V}_2]$ ,  $\bar{\Omega} = \text{blkdiag}(\Omega_1, \Omega_2)$ . The eigenvalue problem (18) is described by

$$\begin{aligned} (\bar{\Omega} \mathbf{R}_T \bar{\Omega} - \bar{\Omega}^2 + \mathbf{I}) \frac{1}{\sqrt{2}} \begin{bmatrix} \mathbf{Q}'_1 & \mathbf{Q}'_1 \\ \mathbf{Q}'_2 & -\mathbf{Q}'_2 \end{bmatrix} &= \begin{bmatrix} \mathbf{I} & \Omega_1 \mathbf{V}_1^\top \mathbf{V}_2 \Omega_2 \\ \Omega_2 \mathbf{V}_2^\top \mathbf{V}_1 \Omega_1 & \mathbf{I} \end{bmatrix} \frac{1}{\sqrt{2}} \begin{bmatrix} \mathbf{Q}'_1 & \mathbf{Q}'_1 \\ \mathbf{Q}'_2 & -\mathbf{Q}'_2 \end{bmatrix} \\ &= \frac{1}{\sqrt{2}} \begin{bmatrix} \mathbf{Q}'_1 + \mathbf{Q}'_1 \Theta_{12} & \mathbf{Q}'_1 - \mathbf{Q}'_1 \Theta_{12} \\ \mathbf{Q}'_2 \Theta_{12} + \mathbf{Q}'_2 & \mathbf{Q}'_2 \Theta_{12} - \mathbf{Q}'_2 \end{bmatrix} = \frac{1}{\sqrt{2}} \begin{bmatrix} \mathbf{Q}'_1 & \mathbf{Q}'_1 \\ \mathbf{Q}'_2 & -\mathbf{Q}'_2 \end{bmatrix} \begin{bmatrix} \mathbf{I} + \Theta_{12} & \mathbf{0} \\ \mathbf{0} & \mathbf{I} - \Theta_{12} \end{bmatrix}, \end{aligned}$$

where we use  $\Omega_1 \mathbf{V}_1^\top \mathbf{V}_2 \Omega_2 = \mathbf{Q}'_1 \Theta_{12} \mathbf{Q}'_2^\top$  (SVD) in gMSM (14). The projection vectors are described by using the eigenvectors of the smaller eigenvalues as  $\mathbf{D} = (\mathbf{V}_1 \Omega_1 \mathbf{Q}'_1 - \mathbf{V}_2 \Omega_2 \mathbf{Q}'_2) (\mathbf{I} - \Theta_{12})^{-\frac{1}{2}}$ , which are differential vectors between the canonical vectors of the two classes in gMSM.

The proposed gCMSM has the following property for the projection vectors.

**Theorem 3.** *The projection vectors in (19) have the norms less than 1.*

*Proof.* The squared norm of the projection vector is

$$\begin{aligned} \mathbf{p}^\top \mathbf{p} &= \sigma^{-\frac{1}{2}} \mathbf{q}^\top \bar{\mathbf{V}}^\top \bar{\mathbf{V}} \mathbf{q} \sigma^{-\frac{1}{2}} = \sigma^{-\frac{1}{2}} \mathbf{q}'^\top \bar{\Omega} \mathbf{R}_T \bar{\Omega} \mathbf{q}' \sigma^{-\frac{1}{2}} \\ &= \sigma^{-\frac{1}{2}} \mathbf{q}'^\top (\bar{\Omega} \mathbf{R}_T \bar{\Omega} - \bar{\Omega}^2 + \mathbf{I}) \mathbf{q}' \sigma^{-\frac{1}{2}} - \sigma^{-\frac{1}{2}} \mathbf{q}'^\top (\mathbf{I} - \bar{\Omega}^2) \mathbf{q}' \sigma^{-\frac{1}{2}} \quad (20) \end{aligned}$$

$$\leq \sigma^{-\frac{1}{2}} \mathbf{q}'^\top \sigma \mathbf{q}' \sigma^{-\frac{1}{2}} = \mathbf{q}'^\top \mathbf{q}' = 1, \quad (21)$$

where we use  $\mathbf{I} - \bar{\Omega}^2 \succcurlyeq 0$  since the weights are  $\omega_l \leq 1, \forall l$ , and the eigenvalue problem (18) for transforming (20) to (21), and  $\|\mathbf{q}'\| = 1$  to get the last equality.  $\square$

### 3.3 gCMSM for multiple subspaces

The gCMSM is also extended so as to cope with multiple subspaces in each class. According to (7) and (16), the optimization problem is formulated by

$$J = \min_{\mathbf{q}} \frac{\sum_{c, c' | c \neq c'}^C \sum_i^{n_c} \sum_j^{n_{c'}} \mathbf{q}_i^{[c]\top} \mathbf{V}_i^{[c]\top} \mathbf{V}_j^{[c']} \mathbf{q}_j^{[c]}}{\sum_c^C \sum_{i, j | i \neq j}^{n_c} \mathbf{q}_i^{[c]\top} \mathbf{V}_i^{[c]\top} \mathbf{V}_j^{[c]} \mathbf{q}_j^{[c]} + \mathbf{q}^\top \bar{\Omega}^{-2} \mathbf{q}} \quad (22)$$

$$\begin{aligned} &= \min_{\mathbf{q}} \frac{\mathbf{q}^\top \mathbf{R}_B \mathbf{q}}{\mathbf{q}^\top (\mathbf{R}_W - \mathbf{I} + \bar{\Omega}^{-2}) \mathbf{q}} \Rightarrow \mathbf{R}_B \mathbf{Q} = (\mathbf{R}_W - \mathbf{I} + \bar{\Omega}^{-2}) \mathbf{Q} \bar{\Sigma} \\ &\Leftrightarrow (\mathbf{R}_T - \mathbf{I} + \bar{\Omega}^{-2}) \mathbf{Q} = (\mathbf{R}_W - \mathbf{I} + \bar{\Omega}^{-2}) \mathbf{Q} \Sigma \quad (23) \end{aligned}$$

$$\Leftrightarrow (\bar{\Omega} \mathbf{R}_T \bar{\Omega} - \bar{\Omega}^2 + \mathbf{I}) \mathbf{Q}' = (\bar{\Omega} \mathbf{R}_W \bar{\Omega} - \bar{\Omega}^2 + \mathbf{I}) \mathbf{Q}' \Sigma, \quad (24)$$



where  $\Sigma = \tilde{\Sigma} + \mathbf{I}$ ,  $\mathbf{Q} = \bar{\Omega}\mathbf{Q}'$  and

$$\begin{aligned}\bar{\mathbf{V}}^{[c]} &= [\mathbf{V}_1^{[c]}, \dots, \mathbf{V}_{n_c}^{[c]}], \quad \bar{\mathbf{V}} = [\bar{\mathbf{V}}^{[1]}, \dots, \bar{\mathbf{V}}^{[C]}], \\ \bar{\Omega}^{[c]} &= \text{blkdiag}(\{\Omega_i^{[c]}\}_{i=1}^{n_c}), \quad \bar{\Omega} = \text{blkdiag}(\{\bar{\Omega}^{[c]}\}_{c=1}^C), \\ \mathbf{R}_W &= \text{blkdiag}(\{\bar{\mathbf{V}}^{[c]\top} \bar{\mathbf{V}}^{[c]}\}_{c=1}^C) \succcurlyeq 0, \quad \mathbf{R}_T = \bar{\mathbf{V}}^\top \bar{\mathbf{V}} \succcurlyeq 0, \quad \mathbf{R}_B = \mathbf{R}_T - \mathbf{R}_W.\end{aligned}$$

The projection vectors are given by using the eigenvectors of the smaller eigenvalues as

$$\mathbf{D} = \bar{\mathbf{V}}\mathbf{Q}\Sigma^{-\frac{1}{2}} = \bar{\mathbf{V}}\bar{\Omega}\mathbf{Q}'\Sigma^{-\frac{1}{2}}. \quad (25)$$

The generalized eigenvalue problem (24) is transformed to a standard eigenvalue problem as follows. We apply an eigen-decomposition to  $\bar{\Omega}^{[c]}\bar{\mathbf{V}}^{[c]\top}\bar{\mathbf{V}}^{[c]}\bar{\Omega}^{[c]} - \bar{\Omega}^{[c]2} + \mathbf{I} = \mathbf{Q}'^{[c]}\Sigma^{[c]}\mathbf{Q}'^{[c]\top}$ , which is *within-class* gCMSM, and define  $\mathbf{R}_W'^{\frac{1}{2}} = \text{blkdiag}(\{\mathbf{Q}'^{[c]}\Sigma^{[c]^{-\frac{1}{2}}}\}_{c=1}^C)$  such that  $\mathbf{R}_W'^{\frac{1}{2}\top}(\bar{\Omega}\mathbf{R}_W\bar{\Omega} - \bar{\Omega}^2 + \mathbf{I})\mathbf{R}_W'^{\frac{1}{2}} = \mathbf{I}$ . Therefore, (24) is transformed into

$$\mathbf{R}_W'^{\frac{1}{2}\top}(\bar{\Omega}\mathbf{R}_T\bar{\Omega} - \bar{\Omega}^2 + \mathbf{I})\mathbf{R}_W'^{\frac{1}{2}}\mathbf{Q}'' = \mathbf{Q}''\Sigma, \quad \text{s.t.}, \quad \mathbf{Q}''^\top\mathbf{Q}'' = \mathbf{I}, \quad \text{where } \mathbf{Q}' = \mathbf{R}_W'^{\frac{1}{2}}\mathbf{Q}'' \quad (26)$$

The left-hand-side matrix is

$$\begin{aligned}\mathbf{R}_W'^{\frac{1}{2}\top}(\bar{\Omega}\mathbf{R}_T\bar{\Omega} - \bar{\Omega}^2 + \mathbf{I})\mathbf{R}_W'^{\frac{1}{2}} \\ = [\mathbf{P}^{[1]}, \dots, \mathbf{P}^{[C]}]^\top [\mathbf{P}^{[1]}, \dots, \mathbf{P}^{[C]}] - \text{blkdiag}(\{\mathbf{P}^{[c]\top}\mathbf{P}^{[c]}\}_{c=1}^C) + \mathbf{I},\end{aligned}$$

where  $\mathbf{P}^{[c]} = \bar{\mathbf{V}}^{[c]}\bar{\Omega}^{[c]}\mathbf{Q}'^{[c]}\Sigma^{[c]^{-\frac{1}{2}}}$  indicates the projection vectors in the  $c$ -th within-class gCMSM in (19), and the final projection vectors (25) are described by  $\mathbf{D} = \bar{\mathbf{V}}\bar{\Omega}\mathbf{R}_W'^{\frac{1}{2}}\mathbf{Q}''\Sigma^{-\frac{1}{2}} = [\mathbf{P}^{[1]}, \dots, \mathbf{P}^{[C]}\mathbf{Q}''\Sigma^{-\frac{1}{2}}$ . Therefore, (26) is regarded as *between-class* gCMSM using the vectors produced by *within-class* gCMSM. The procedure for this method is shown in Algorithm 1 and the practical details are described in the next section.

This proposed method has the following properties regarding the projection vectors  $\mathbf{D}$  and the eigenvalues  $\Sigma = \text{diag}(\{\sigma_l\}_l)$ .

**Theorem 4.** *The projection vectors in (25) have the norms less than 1.*

*Proof.* The squared norm of the projection vector is

$$\begin{aligned}\mathbf{d}^\top\mathbf{d} &= \sigma^{-\frac{1}{2}}\mathbf{q}'^\top\bar{\Omega}\bar{\mathbf{V}}^\top\bar{\mathbf{V}}\bar{\Omega}\mathbf{q}'\sigma^{-\frac{1}{2}} \\ &= \sigma^{-\frac{1}{2}}\mathbf{q}'^\top(\bar{\Omega}\mathbf{R}_T\bar{\Omega} - \bar{\Omega}^2 + \mathbf{I})\mathbf{q}'\sigma^{-\frac{1}{2}} - \sigma^{-\frac{1}{2}}\mathbf{q}'^\top(\mathbf{I} - \bar{\Omega}^2)\mathbf{q}'\sigma^{-\frac{1}{2}} \\ &\leq \sigma^{-\frac{1}{2}}\mathbf{q}'^\top\sigma(\bar{\Omega}\mathbf{R}_W\bar{\Omega} - \bar{\Omega}^2 + \mathbf{I})\mathbf{q}'\sigma^{-\frac{1}{2}} = \mathbf{q}''^\top\mathbf{q}'' = 1,\end{aligned}$$

where we use  $\mathbf{I} - \bar{\Omega}^2 \succcurlyeq 0$  and the eigenvalue problems (24) and (26).  $\square$

**Theorem 5.** *The eigenvalues in (24) are bounded in  $0 \leq \sigma_l \leq C$ ,  $\forall l$ .*

*Proof.* The weights  $\omega \leq 1$  result in  $\bar{\Omega}^{-2} - \mathbf{I} \succcurlyeq 0$ . Thus, since  $\mathbf{R}_T - \mathbf{I} + \bar{\Omega}^{-2} \succcurlyeq 0$  and  $\mathbf{R}_W - \mathbf{I} + \bar{\Omega}^{-2} \succcurlyeq 0$ , we get  $\sigma_l \geq 0$ ,  $\forall l$  in (23). By using  $\mathbf{R}_W - \mathbf{I} + \bar{\Omega}^{-2} \succcurlyeq \mathbf{R}_W$ , the eigenvalues are upper-bounded by

$$\max_l \sigma_l = \max_q \frac{\mathbf{q}^\top \mathbf{R}_B \mathbf{q}}{\mathbf{q}^\top (\mathbf{R}_W - \mathbf{I} + \bar{\Omega}^{-2}) \mathbf{q}} + 1 \leq \max_q \frac{\mathbf{q}^\top \mathbf{R}_B \mathbf{q}}{\mathbf{q}^\top \mathbf{R}_W \mathbf{q}} + 1 = \max_q \frac{\mathbf{q}^\top \mathbf{R}_T \mathbf{q}}{\mathbf{q}^\top \mathbf{R}_W \mathbf{q}} \leq C,$$

**Algorithm 1** : gCMSM training

---

**Input:**  $\mathbf{X}_i^{[c]} \in \mathbb{R}^{d \times n_{ci}}$ :  $d$ -dimensional feature vectors of the  $i$ -th set in the  $c$ -th class.

- 1: **for**  $c = 1$  to  $C$  **do**
- 2: Subspace:  $\mathbf{X}_i^{[c]} \mathbf{X}_i^{[c]\top} = \mathbf{V}_i^{[c]} \text{diag}(\boldsymbol{\lambda}_i^{[c]}) \mathbf{V}_i^{[c]\top}$ ,  $i = 1, \dots, n_c$ .
- 3:  $\bar{\mathbf{V}}^{[c]} = [\mathbf{V}_1^{[c]}, \dots, \mathbf{V}_{n_c}^{[c]}]$ ,  $\bar{\boldsymbol{\Omega}}^{[c]} = \text{diag}([\mathbf{w}_m(\boldsymbol{\lambda}_1^{[c]})^\top, \dots, \mathbf{w}_m(\boldsymbol{\lambda}_{n_c}^{[c]})^\top]^\top)$ .
- 4: *Within-class* gCMSM:  $\bar{\boldsymbol{\Omega}}^{[c]} \bar{\mathbf{V}}^{[c]\top} \bar{\mathbf{V}}^{[c]} \bar{\boldsymbol{\Omega}}^{[c]} - \bar{\boldsymbol{\Omega}}^{[c]2} + \mathbf{I} = \mathbf{Q}^{[c]} \boldsymbol{\Sigma}^{[c]} \mathbf{Q}^{[c]\top}$ ,  
 $\mathbf{P}^{[c]} = \bar{\mathbf{V}}^{[c]} \bar{\boldsymbol{\Omega}}^{[c]} \mathbf{Q}^{[c]} \boldsymbol{\Sigma}^{[c]-\frac{1}{2}}$ .
- 5: Cut-off the vectors:  $\mathbf{P}^{[c]} \leftarrow \{\mathbf{p}_l^{[c]} \mid \|\mathbf{p}_l^{[c]}\|^2 > \epsilon\}$ ,  $\mathbf{p}_l^{[c]}$ : the  $l$ -th column vector in  $\mathbf{P}^{[c]}$ .
- 6: **end for**
- 7:  $\bar{\mathbf{P}} = [\mathbf{P}^{[1]}, \dots, \mathbf{P}^{[C]}]$ ,  $\mathbf{O} = \text{blkdiag}(\{\mathbf{P}^{[c]\top} \mathbf{P}^{[c]}\}_{c=1}^C)$ : block-diagonal matrix.
- 8: *Between-class* gCMSM:  $(\bar{\mathbf{P}}^\top \bar{\mathbf{P}} - \mathbf{O} + \mathbf{I}) \mathbf{Q} = \mathbf{Q} \boldsymbol{\Sigma}$ ,  $\mathbf{D} = \bar{\mathbf{P}} \mathbf{Q} \boldsymbol{\Sigma}^{-\frac{1}{2}}$ .
- 9: Cut-off the projections:  $\mathbf{D} \leftarrow \{\mathbf{d}_l \mid \sigma_l < \tau\}$ ,  $\mathbf{d}_l$ : the  $l$ -th column vector in  $\mathbf{D}$ .

**Output:** Projection vectors:  $\mathbf{D} \in \mathbb{R}^{d \times L}$ .

---

**Algorithm 2** : gMSM classification

---

**Input:**  $\{\mathbf{X}_i, \mathbf{y}_i\}_{i=1}^n$ : pairs of  $d$ -dimensional feature vectors  $\mathbf{X}_i \in \mathbb{R}^{d \times n_i}$  and a class label  $\mathbf{y}_i \in \{1, \dots, C\}$  of the  $i$ -th set for training,  
 $\mathbf{X}_t$ : feature vectors of a set for test.

- 1: Subspace in gCMSM space:  $\mathbf{D}^\top \mathbf{X}_t = \mathbf{V}_t \text{diag}(\boldsymbol{\lambda}_t) \mathbf{U}_t^\top$ ,  $\boldsymbol{\Omega}_t = \text{diag}(\mathbf{w}_m(\boldsymbol{\lambda}_t))$ .
- 2: **for**  $i = 1$  to  $n$  **do**
- 3: Subspace in gCMSM space:  $\mathbf{D}^\top \mathbf{X}_i = \mathbf{V}_i \text{diag}(\boldsymbol{\lambda}_i) \mathbf{U}_i^\top$ ,  $\boldsymbol{\Omega}_i = \text{diag}(\mathbf{w}_m(\boldsymbol{\lambda}_i))$ .
- 4: Similarity:  $s_{ti} = \text{tr}(\boldsymbol{\Theta}_{ti}^2)$ , where  $\boldsymbol{\Omega}_t \mathbf{V}_t^\top \mathbf{V}_i \boldsymbol{\Omega}_i = \mathbf{Q}_t \boldsymbol{\Theta}_{ti} \mathbf{Q}_i^\top$  (SVD).
- 5: **end for**
- 6:  $k$ -NN:  $\mathbf{y} = k\text{NN}(\{s_{ti}, \mathbf{y}_i\}_{i=1}^n)$ .

**Output:** Estimated class label of the test set  $\mathbf{X}_t$ :  $\mathbf{y}$ .

---

where the last inequality is obtained from Theorem 2. And, as in CMSM (Sec.2.3),  $\sigma_l < 1$  indicates  $J < 0$  in (22), producing the discriminative coefficients  $\mathbf{q}$ .  $\square$

**3.4 Procedure for classification**

For classifying sets of feature vectors, we first train by gCMSM the discriminative space in which the gMSM is subsequently performed to compute the subspace angles. The training procedure by gCMSM is shown in Algorithm 1. In the line 5, we cut off the class representative vectors  $\mathbf{P}^{[c]}$  which are fed into *between-class* gCMSM in the line 8. By Theorem 3, the vectors  $\mathbf{P}^{[c]}$  have the norms less than 1, and there are negligible vectors whose norms are close to 0. By eliminating those vectors, the eigenvalue problem in the line 8 is sufficiently speeded up without changing the result; in this paper, we set  $\epsilon = 0.1$ . In the line 9, we extract the discriminative projection vectors of the smaller eigenvalues based on Theorem 5; note that the projection vectors of the eigenvalues less than 1 are discriminative, and we set  $\tau = 0.9999$ .

The procedure to classify sets is shown in Algorithm 2. We apply gMSM in the discriminative space produced by Algorithm 1 and then  $k$ -NN, say  $k = 1$ .

## 4 Experimental results

We conducted the experiments on 3D object classification by using ETH-80 [19] and RGB-D [20] datasets.

### 4.1 ETH-80 dataset

ETH-80 dataset [19] consists of eight categories, each of which contains 10 objects with 41 images of different views. The images were transformed into gray-scale and resized to  $32 \times 32$  pixels after subtracting background as shown in Fig. 3, resulting in 1024-dimensional image vectors. For evaluation, we applied two-fold cross validation; the 10 objects in each category are partitioned into two folds of five objects for training and test. We randomly repeated the validation 10 times and reported the averaged classification performance. Two types of experimental protocol were employed; ‘experiment 1’ is to use all of 41 views and ‘experiment 2’ is to randomly pick up 15 views for test while using all views in training.

First, we validated the theorems that the proposed method (Algorithm 1) is based on. For Theorem 3 used in the line 5 of Algorithm 1, the norms of the class representative vectors  $\mathbf{p}_l^{[c]}$  produced by within-class gCMSM are shown in Fig. 4a where those vectors are sorted in descending order of the eigenvalues  $\sigma_l^{[c]}$ . Some of those vectors have tiny norms and by eliminating those tiny vectors, the computational cost in the subsequent between-class gCMSM is significantly reduced; actually the computation time for gCMSM is reduced to 0.19 sec from 2.13 sec on Xeon 3.33GHz. For Theorem 5 in the line 9 of Algorithm 1, the eigenvalues  $\sigma_l$  of gCMSM are shown in Fig. 4b along with the norms of the projection vectors  $\mathbf{d}_l$  to validate Theorem 4. We can see that around  $\sigma_l = 1$  the projection vectors have smaller norms and are less contributive to the projections; we exploit the projections of  $\sigma_l < \tau = 0.9999$ .

We then analyzed the performances of the proposed methods; gMSM in (13) measures canonical angles in the original vector space while the methods of gCMSM in (18, 24) perform in the discriminative space. Note that there are two types of gCMSM; gCMSM in (18), denoted by gCMSM-mono, groups all the vectors of each class into a single set which is further represented by a mono-subspace, while gCMSM in (24) deals with each object’s image set individually. Fig. 5 shows the performances on various  $m$  used in the weighting function  $\mathbf{w}_m$  in (15). The gCMSM produces the best performance, demonstrating that 1) the discriminative space contributes to improve the performances in comparison to gMSM and 2) the optimization considering *within-class* measures as well as *between-class* ones is effective compared to gCMSM-mono. The proposed methods are also compared to the other subspace-based methods, MSM [7], OSM [10], DCC [5], CMSM [8], as shown in Fig. 5 and Table 1. For fair comparison, the subspace dimensionality is set by the number  $m$  in the weighting function for gCMSM. As is the case with gCMSM, CMSM is superior to CMSM-mono. The proposed gCMSM produces stably high performances on various  $m$  due to soft weighting, which are superior to the others. In the experiment 2, we used only a

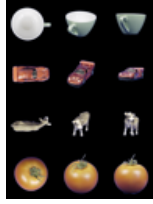
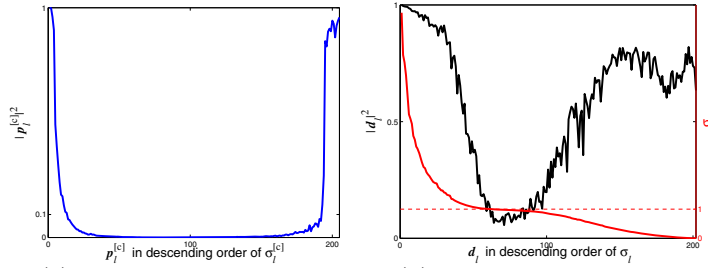
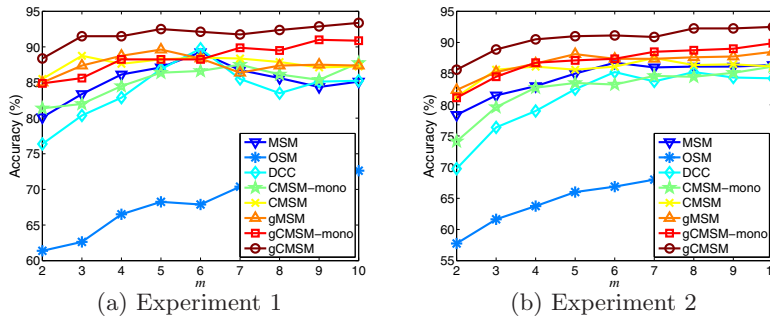


Fig. 3. ETH80



(a) Within-class gCMSM (b) Between-class gCMSM

Fig. 4. Norms of projection vectors with eigenvalues.



(a) Experiment 1

(b) Experiment 2

Fig. 5. ETH-80 dataset.

subset of images of objects in test, making the test set different from the training ones. Even though the principal directions of the image sets are highly sensitive to the variations in pose, the proposed gCMSM exhibits superior performance which is almost the same as in the experiment 1, while the others are degraded.

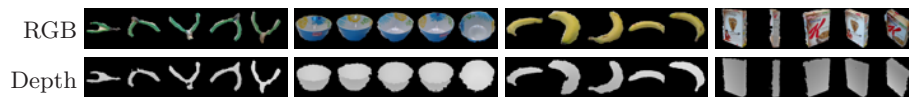
## 4.2 RGB-D dataset

RGB-D dataset [20] contains color and depth images of 300 physically distinct objects in 51 categories. This dataset is composed of video sequences for each object as it is spun around on a turntable at constant speed. The video data is recorded by the camera mounted at three different views of approximately 30, 45 and 60 degrees from the horizon with 20 Hz providing about 250 RGB+depth frames in a sequence. For this experiment, we subsampled the video sequences at every fifth frame to obtain about 40,000 RGB-depth image pairs. The object regions were extracted by background subtraction based on the depth and then resized into  $32 \times 32$  pixels both for RGB and depth images, as shown in Fig. 6. Thus, we obtain the depth features of 1024 dimensionality as well as the gray-scale features transformed from RGB.

The proposed methods are compared to the other methods by using each of gray-scale and depth features. The results are shown in Fig. 7 and Table 2, demonstrating the proposed gCMSM produces stably high performances compared to the others; the best accuracies are competitive even with the state-of-the-art results [20, 21]. The proposed methods are more effectively applied to

**Table 1.** Best accuracies (%) on ETH-80 dataset.

Method	MSM	OSM	DCC	CMSM	gMSM	gCMSM
Exp.1	89.25	72.63	89.75	88.75	89.63	<b>93.38</b>
Exp.2	86.75	70.13	85.25	87.50	88.50	<b>92.50</b>

**Fig. 6.** RGB-D dataset.

the depth features than to the image features, which is contrast to [20, 21] using sophisticated local features such as SIFT. The depth values depending on the object’s distance give less effective clues for classification in the local features, while the global structure (distribution) of the depth values is effectively extracted by the subspace. effective for classification. On the other hand, the appearance (gray-scale) features are affected by the textures and slightly degrade the performance.

## 5 Conclusion

We have proposed generalized mutual subspace based methods. A *soft weighting* is introduced across the bases of the subspace instead of definitely picking up only principal basis vectors. The soft weights are constructed according to the variances, *i.e.*, eigenvalues, associated with the bases. The generalized MSM is proposed to effectively combine the bases via the soft weights for measuring the subspace angles. In addition, CMSM is reformulated with theoretical justification, and the reformulation also enables us to generalize the CMSM by incorporating the soft weighting for providing the discriminative space. In the experiments on 3D object recognition using ETH-80 and RGB-D datasets, the proposed methods exhibited stably favorable performances compared to the other subspace-based methods.

## References

1. R.Wang, Shan, S., Chen, X., Gao, W.: Manifold-manifold distance with application to face recognition based on image set. In: CVPR. (2008)
2. T.Wang, Shi, P.: Kernel grassmannian distances and discriminant analysis for face recognition from image sets. Pattern Recognition Letters **30** (2009) 1161–1165
3. Shakhnarovich, G., Fisher, J.W., Darrel, T.: Face recognition from long-term observations. In: ECCV. (2002) 851–868
4. Arandjelović, O., Shakhnarovich, G., Fisher, J., Cipolla, R., Darrell, T.: Face recognition with image sets using manifold density divergence. In: CVPR. (2005) 581–588
5. Kim, T.K., Kittler, J.V., Cipolla, R.: Discriminative learning and recognition of image set classes using canonical correlations. IEEE Transaction on Pattern Analysis and Machine Intelligence **29** (2007) 1005–1018

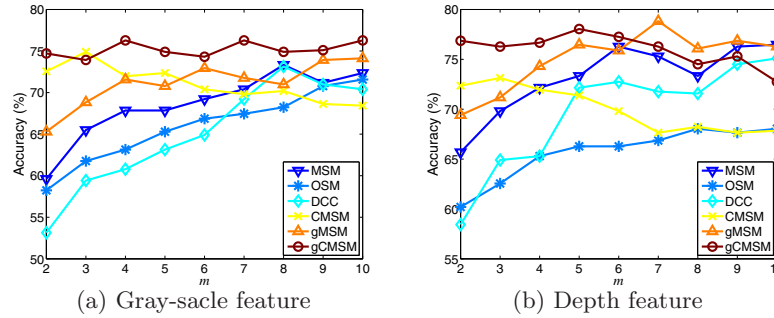


Fig. 7. RGB-D dataset.

Table 2. Best accuracies (%) on RGB-D dataset.

Method	MSM	CMSM	OSM	DCC	gMSM	gCMSM	[20]	[21]
Gray-scale	73.33	74.90	71.57	73.14	74.12	<b>76.27</b>	74.7	76.1
Depth	76.47	73.14	68.04	75.10	<b>78.82</b>	<b>78.04</b>	66.8	75.7

6. Wu, J., Fukui, K.: Multiple view based 3d object classification using ensemble learning of local subspaces. In: ICPR. (2008)
7. Yamaguchi, O., Fukui, K., Maeda, K.: Face recognition using temporal image sequence. In: FG. (1998) 318–323
8. Fukui, K., Yamaguchi, O.: Face recognition using multi-viewpoint patterns for robot vision. In: International Symposium of Robotics Research. (2003) 192–201
9. Fukui, K., Yamaguchi, O.: Constrained mutual subspace method using a generalized difference subspace. IEICE Trans. on Info. & Syst. **87D** (2004) 1622–1631
10. Oja, E.: Subspace Methods for Pattern Recognition. Research Studies Press (1983)
11. Kim, T.K., Kittler, J.V., Cipolla, R.: On-line learning of mutually orthogonal subspaces for face recognition by image sets. IEEE Transaction on Image Processing **19** (2010) 1067–1074
12. Kim, T.K., Arandjelović, O., Cipolla, R.: Learning over sets using boosted manifold principal angles (bompa). In: BMVC. (2005) 779–788
13. Wolf, L., Shashua, A.: Learning over sets using kernel principal angles. Journal of Machine Learning Research **4** (2003) 913–931
14. Watanabe, S., Pakvasa, N.: Subspace method of pattern recognition. In: ICPR. (1973) 25–32
15. Björck, Å., Golub, G.H.: Numerical methods for computing angles between linear subspaces. Mathematics of Computation **27** (1973) 579–594
16. Hotelling, H.: Relations between two sets of variates. Biometrika **28** (1936) 321–372
17. Duda, R.O., Hart, P.E., Stork, D.G.: Pattern Classification. 2 edn. Wiley-Interscience (2001)
18. Vapnik, V.: Statistical Learning Theory. Wiley (1998)
19. Leibe, B., Schiele, B.: Analyzing appearance and contour based methods for object categorization. In: CVPR. (2003) 409–415
20. Lai, K., Bo, L., Ren, X., Fox, D.: A large-scale hierarchical multi-view rgb-d object dataset. In: ICRA. (2011) 1817–1824
21. Bo, L., Lai, K., Ren, X., Fox, D.: Object recognition with hierarchical kernel descriptors. In: CVPR. (2011) 1729–1736