Anadolu Üniversitesi Bilim ve Teknoloji Dergisi A- Uygulamalı Bilimler ve Mühendislik
Anadolu University Journal of Science and Technology A- Applied Sciences and Engineering

2016 - Volume: 17 Number: 2
Page: 251 - 262
DOI: 10.18038/btda.68101
Received: 24 March 2016 Revised: 28 April 2016 Accepted: 02 May 2016

A NOVEL IMPLEMENTATION ALGORITHM FOR CALCULATION OF COMMON VECTORS

Mehmet KOÇ[1, *], Atalay BARKANA[2]

[1]Department of Electrical and Electronics Engineering, Engineering Faculty, Bilecik Seyh Edebali University, 11210, Bilecik, TURKEY

[2]Department of Electrical and Electronics Engineering, Engineering Faculty, Anadolu University, 26555, Eskişehir, TURKEY

ABSTRACT

Common vector approach (CVA), discriminative common vector approach (DCVA), and linear regression classification (LRC) are subspace methods used in pattern recognition. Up to now, there were two well-known algorithms to calculate the common vectors: (*i*) by using the Gram-Schmidt orthogonalization process, (*ii*) by using the within-class covariance matrices. The purpose of this paper is to introduce a new implementation algorithm for the derivation of the common vectors using the linear regression idea. The derivation of the discriminative common vectors through LRC is also included in this paper. Two numerical examples are given to clarify the proposed derivations. An experimental work is given in AR face database to compare the recognition performances of CVA, DCVA, and LRC. Additionally, the three implementation algorithms of common vector are compared in terms of processing time efficiency.

Keywords: Common vector, Discriminative common vector, Linear regression classification, Subspace methods, Face recognition

1. INTRODUCTION

The dimension of the feature space is very important in face recognition (FR) problems. An *n*-by-*m* sized face image corresponds to a point in *nm*-dimensional space. This situation raises some difficulties such as computational cost, increasing measurement and storage requirements, curse of dimensionality, etc. [1]. Many subspace methods are proposed to reduce dimensionality and increase the recognition performance of FR systems [1-5].

In this work, we propose a new algorithm to obtain the common vectors using linear regression classification (LRC) method [4]. We provide a numerical example to show the equivalence of the proposed derivation with the previous two derivations of the common vectors [2,6]. We also express a new derivation of the discriminative common vectors using LRC. In this section, brief reviews of the common vector approach (CVA), discriminative common vector approach (DCVA) [3], and linear regression classification (LRC) are given. Derivation of common vector (CV) using LRC approach and a numerical example are given in Section 2. Derivation of discriminative common vector (DCV) using LRC approach with a numerical example is given in Section 3. A real life application of face recognition is given in Section 4 and the conclusion of the work is given in Section 5.

1.1. Common Vector Approach

There are two basic algorithms to calculate the common vector of a class [2,6]. The first approach is based on Gram-Schmidt orthogonalization and the other is based on covariance matrix of a class.

*Corresponding Author: : mehmet.koc@bilecik.edu.tr

(*i*) *Gram-Schmidt Orthogonalization Approach*: Let $\{\boldsymbol{a}_1^i, \boldsymbol{a}_2^i, \dots, \boldsymbol{a}_m^i\}$ be the *n*-dimensional feature vectors used in the i^{th} class training set. And let the total number of classes be C, that is, $i = 1,2, \dots, C$. The difference vectors for each class are obtained as

$$\begin{aligned} \boldsymbol{b}_1^i &= \boldsymbol{a}_2^i - \boldsymbol{a}_1^i, \\ \boldsymbol{b}_2^i &= \boldsymbol{a}_3^i - \boldsymbol{a}_1^i, \\ &\vdots \\ \boldsymbol{b}_{m-1}^i &= \boldsymbol{a}_m^i - \boldsymbol{a}_1^i. \end{aligned} \tag{1}$$

The orthonormal vector set $\boldsymbol{Z}_i = [\boldsymbol{z}_1 \vdots \boldsymbol{z}_2 \vdots \cdots \vdots \boldsymbol{z}_{m-1}]$ that spans the difference subspace of the i^{th} class can be obtained by applying the Gram-Schmidt orthogonalization procedure to the difference vectors given in Eq.(1). Then the common vector of the i^{th} class can be obtained by subtracting the projection of any feature vector onto the difference subspace from itself as it is shown below

$$\boldsymbol{a}_{com}^i = \boldsymbol{a}_j^i - \sum_{k=1}^{m-1} (\boldsymbol{z}_k^T \boldsymbol{a}_j^i) \boldsymbol{z}_k \ , i = 1, \dots, C, j = 1, \dots, m \tag{2}$$

Due to the above formula, common vector is considered to be the projection of any feature vector in a class onto the indifference subspace. The indifference subspace in here is the complementary subspace of the difference subspace. That is, let the projection matrix onto the difference subspace be the matrix $\boldsymbol{P}$ where $\boldsymbol{P} = \sum_{k=1}^{m-1} \boldsymbol{z}_k \boldsymbol{z}_k^T$. Then the projection matrix onto the indifference subspace, $\boldsymbol{P}^{\perp}$, will be found from $\boldsymbol{P} + \boldsymbol{P}^{\perp} = \boldsymbol{I}$.

(*ii*) *Covariance Matrix Approach*: The within-class covariance matrix is used to calculate the common vector of a class. Let the within-class covariance matrix be defined as $\boldsymbol{\Phi}_i = \boldsymbol{A}_i \boldsymbol{A}_i^T$ where $\boldsymbol{A}_i$ is a matrix of the form

$$\boldsymbol{A}_i = [\boldsymbol{a}_1^i - \boldsymbol{\mu}_i \vdots \boldsymbol{a}_2^i - \boldsymbol{\mu}_i \vdots \cdots \vdots \boldsymbol{a}_m^i - \boldsymbol{\mu}_i], \qquad i = 1, \dots, C \tag{3}$$

and $\boldsymbol{\mu}_i$ is the mean of the i^{th} class. Let the pair $(\lambda_k, \boldsymbol{u}_k), i = m, \dots, n$ be the zero eigenvalue and corresponding eigenvector set. Then the projection matrix onto the indifference subspace of the i^{th} class is $\boldsymbol{P}_i^{\perp} = \sum_{k=m}^{n} \boldsymbol{u}_k \boldsymbol{u}_k^T$ where $\boldsymbol{u}_k$s are the eigenvectors correspond to the zero eigenvalues. Then the common vector of the i^{th} class is calculated from

$$\boldsymbol{a}_{com}^i = \boldsymbol{P}_i^{\perp} \boldsymbol{a}_j^i, \qquad i = 1, \dots, C, \qquad j = 1, \dots, m. \tag{4}$$

1.2. Discriminative Common Vector Approach

Discriminative common vector approach is a popular method in face recognition area. Several method related with DCVA have been published [7, 10]. In DCVA, similar with CVA, common vectors can be calculated with Gram-Schmidt orthogonalization process or by using the within-class covariance matrices.

(*i*) *Gram-Schmidt Orthogonalization Approach*: Let the difference vectors of the i^{th} class $\boldsymbol{b}_j^i = \boldsymbol{a}_j^i - \boldsymbol{a}_1^i,\ j = 2, \dots, m$. The orthonormal vector set $\{\boldsymbol{z}_i\}, i = 1,2, \dots, C(m-1)$ is obtained by applying the Gram-Schmidt orthogonalization process to the difference vectors $\boldsymbol{b}_j^i$. Then the common vector of the i^{th} class is calculated by subtracting the projection of any feature vector from the i^{th} class onto the orthonormal vector set $\{\boldsymbol{z}_i\}, i = 1,2, \dots, C(m-1)$ as below:

$$\boldsymbol{a}_{com}^{i} = \boldsymbol{a}_{j}^{i} - \sum_{k=1}^{C(m-1)} \left(\boldsymbol{z}_{k}^{T}\boldsymbol{a}_{j}^{i}\right)\boldsymbol{z}_{k}\,, j = 1, \dots, m \tag{5}$$

(*ii*) *Covariance Matrix Approach*: Different from CVA, the common vectors are calculated using the total within-class scatter matrix, $\boldsymbol{\Phi}_{T} = \boldsymbol{A}\boldsymbol{A}^{T}$, where

$$\boldsymbol{A} = \left[\boldsymbol{a}_{1}^{1} - \boldsymbol{\mu}_{1} \vdots \cdots \vdots \boldsymbol{a}_{m}^{1} - \boldsymbol{\mu}_{1} \vdots \boldsymbol{a}_{1}^{2} - \boldsymbol{\mu}_{2} \vdots \cdots \vdots \boldsymbol{a}_{1}^{C} - \boldsymbol{\mu}_{C}\right] \tag{6}$$

or else $\boldsymbol{\Phi}_{T} = \boldsymbol{\Phi}_{1} + \boldsymbol{\Phi}_{2} + \cdots + \boldsymbol{\Phi}_{C}$.

After eigen-decomposition, the eigenvectors $\{\boldsymbol{u}_{i}\}, i = 1,2, \dots, C(m-1)$ corresponding to the nonzero eigenvalues form the projection matrix $\boldsymbol{P} = \sum_{i=1}^{C(m-1)} \boldsymbol{u}_{i}\boldsymbol{u}_{i}^{T}$ onto the difference subspace of $\boldsymbol{\Phi_{T}}$. Then the common vectors are calculated as in Eq.(4).

After the calculation of the common vectors in DCVA, a final attempt must be made for finding what is called as the discriminant common vectors (DCVs) [**3**]. DCVs are obtained by using the subspace spanned by the common vectors in Eq.(5). For this reason, the difference between the common vectors should be calculated first. Then the difference vectors that span the subspace of the difference of the common vectors can be written as

$$\boldsymbol{b}_{com}^{i} = \boldsymbol{a}_{com}^{i+1} - \boldsymbol{a}_{com}^{1}, i = 1, \dots, C-1, \tag{7}$$

where $\boldsymbol{a}_{com}^{1}$ is the common vector of the first class. It can also be the common vector of any other class. Then the column vectors of the matrix $\boldsymbol{U}$ are the orthonormal basis vector set $\{\boldsymbol{u}_{1}, \boldsymbol{u}_{2}, \dots, \boldsymbol{u}_{C-1}\}$ that can be obtained by applying the Gram-Schmidt orthogonalization procedure to the difference vectors between the commons from Eq. (7), that is, $\boldsymbol{U} = [\boldsymbol{u}_{1} \vdots \boldsymbol{u}_{2} \vdots \cdots \vdots \boldsymbol{u}_{C-1}]$. As a second method, the projection matrix $\boldsymbol{U}$ can be calculated by simply applying the PCA to the covariance matrix of the common vectors. The matrix $\boldsymbol{U}$ in this case is a projection matrix from an *n*-dimensional space to $C-1$ dimensional subspace defined by the difference of the common vectors. Finally, the discriminative common vector of a class is calculated as

$$\boldsymbol{\Omega}_{i} = \boldsymbol{U}^{T}\boldsymbol{a}_{i}^{j}, i = 1, \dots, C, \quad j = 1, \dots, m. \tag{8}$$

Here, just to emphasize the result, all the vectors are projected onto the $C-1$ dimensional subspace formed by an orthogonal base of the differences of all the common vectors.

1.3. Linear Regression Classification

Linear regression classification (LRC) has become very popular lately. Several methods are proposed which are inspired by LRC [11-13]. LRC is generally used for face recognition problems. Authors in [14]. propose an improved version of LRC which can be applicable to low dimensional datasets.

Let *C* be the number of classes, *m* be the number of feature vectors of a class used in training, and let $\{\boldsymbol{a}_{1}^{i}, \boldsymbol{a}_{2}^{i}, \dots, \boldsymbol{a}_{m}^{i}\}$ be the feature vectors of the i^{th} class training set as it was in the previous section. The set of these feature vectors spans a subspace belonging to the i^{th} class. LRC idea is based on a distance metric given by

$$d_{i} = \left|\boldsymbol{a}_{x} - \boldsymbol{a}_{x}^{i}\right| \tag{9}$$

where $\boldsymbol{a}_x$ is the test feature vector and $\boldsymbol{a}_x^i$ is its projection onto the subspace spanned by the training set of the *i*^th^ class. Let $\boldsymbol{P}_i$ be the projection matrix onto the subspace spanned by the feature vectors of the *i*^th^ class, then the distance metric will become

$$d_i = |\boldsymbol{a}_x - \boldsymbol{P}_i \boldsymbol{a}_x|. \tag{10}$$

In LRC method, $\boldsymbol{P}_i$ is calculated using a linear combination of the feature vectors in the i^{th} class the training set under a constraint relation in its optimized form in terms of minimum sum of error squares. Let $\boldsymbol{a}_x^i = \boldsymbol{W}_i \boldsymbol{\beta}_i$ be the estimation of $\boldsymbol{a}_x$ where $\boldsymbol{W}_i$ is a matrix formed from the feature vectors in the i^{th} class the training set, i.e.,

$$\boldsymbol{W}_i = \left[\boldsymbol{a}_1^i \vdots \boldsymbol{a}_2^i \vdots \cdots \vdots \boldsymbol{a}_m^i\right] \tag{11}$$

The projection of a feature vector onto the i^{th} class subspace can be calculated from

$$\boldsymbol{a}_x = \boldsymbol{W}_i \boldsymbol{\beta}_i + \boldsymbol{\varepsilon} \tag{12}$$

where $\boldsymbol{\varepsilon}$ is the error or the remaining part of $\boldsymbol{a}_x$ in the rest of the whole space $\mathbb{R}^n$ of feature vectors. The sum of the error squares can be formed with ease

$$S = \boldsymbol{\varepsilon}^T \boldsymbol{\varepsilon} = (\boldsymbol{a}_x - \boldsymbol{W}_i \boldsymbol{\beta}_i)^T (\boldsymbol{a}_x - \boldsymbol{W}_i \boldsymbol{\beta}_i). \tag{13}$$

To minimize *S*, the critical point(s) must be calculated by taking the derivative of *S* with respect to $\boldsymbol{\beta}_i$.

$$\begin{gathered} \frac{\partial S}{\partial \boldsymbol{\beta}_i} = \boldsymbol{0} \\ \frac{\partial}{\partial \boldsymbol{\beta}_i}\left[\boldsymbol{a}_x^T \boldsymbol{a}_x - \boldsymbol{\beta}_i^T \boldsymbol{W}_i^T \boldsymbol{a}_x - \boldsymbol{a}_x^T \boldsymbol{W}_i \boldsymbol{\beta}_i + \boldsymbol{\beta}_i^T \boldsymbol{W}_i^T \boldsymbol{W}_i \boldsymbol{\beta}_i\right] = \boldsymbol{0} \\ -2\boldsymbol{W}_i^T \boldsymbol{a}_x + 2\boldsymbol{W}_i^T \boldsymbol{W}_i \boldsymbol{\beta}_i = \boldsymbol{0} \\ \boldsymbol{\beta}_i = \left(\boldsymbol{W}_i^T \boldsymbol{W}_i\right)^{-1} \boldsymbol{W}_i^T \boldsymbol{a}_x \end{gathered} \tag{14}$$

Then the distance metric of LRC becomes

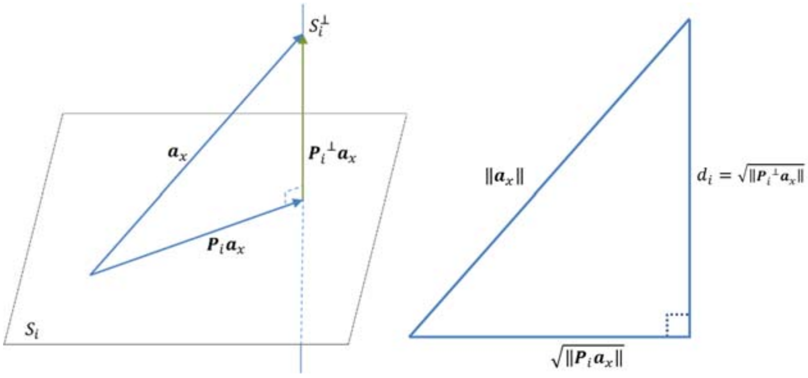
$$d_i = \left|\boldsymbol{a}_x - \boldsymbol{a}_x^i\right| = |\boldsymbol{a}_x - \boldsymbol{W}_i \boldsymbol{\beta}_i| = \left|\boldsymbol{a}_x - \boldsymbol{W}_i \left(\boldsymbol{W}_i^T \boldsymbol{W}_i\right)^{-1} \boldsymbol{W}_i^T \boldsymbol{a}_x\right|. \tag{15}$$

Also, by comparing Eq.(10) with Eq.(15), the projection matrix onto the subspace determined by the i^{th} class $\boldsymbol{P}_i = \boldsymbol{W}_i \left(\boldsymbol{W}_i^T \boldsymbol{W}_i\right)^{-1} \boldsymbol{W}_i^T$ is obtained. If the orthogonal complement of the projection matrix $\boldsymbol{P}_i$ is shown with $\boldsymbol{P}_i^{\perp}$, then the metric given in Eq.(15) is equal to

$$d_i = \left|\boldsymbol{P}_i^{\perp} \boldsymbol{a}_x\right|. \tag{16}$$

Here $\boldsymbol{P}_i^{\perp} \boldsymbol{a}_x$ is the projection of $\boldsymbol{a}_x$ onto a subspace that complements the subspace of the i^{th} class to the whole space with the relation $\boldsymbol{P}_i + \boldsymbol{P}_i^{\perp} = \boldsymbol{I}$. Therefore d_i is the distance of $\boldsymbol{a}_x$ to the subspace formed from the feature vectors of the i^{th} class. In that sense if $\boldsymbol{a}_x$ belongs to the i^{th} class, $\boldsymbol{P}_i^{\perp} \boldsymbol{a}_x$ must be negligibly small, or $\boldsymbol{P}_i \boldsymbol{a}_x$ is almost equal to $\boldsymbol{a}_x$. The geometric illustration of the classification is shown in Figure 1. Here S_i and $S_i^{\perp}$ are the subspaces spanned by the training vectors set of the i^{th} class and its complementary subspace respectively. As d_i gets smaller the similarity between $\boldsymbol{P}_i \boldsymbol{a}_x$ and $\boldsymbol{a}_x$ increases.

Also with the above relations it is known that if $\boldsymbol{a}_x \epsilon \{\boldsymbol{a}_1, \boldsymbol{a}_2, \dots, \boldsymbol{a}_m\}$ or if $\boldsymbol{a}_x = \alpha_1 \boldsymbol{a}_1 + \alpha_2 \boldsymbol{a}_2 + \dots + \alpha_m \boldsymbol{a}_m$, where $\alpha_i \in \mathbb{R}$, or in other words if $\boldsymbol{a}_x$ is within the subspace of the i^{th} class, then $d_i = 0$.



(a) (b)

Figure 1. The geometric illustration of the relation between the components of $\boldsymbol{a}_x$. The feature space is separated into two complementary subspaces by using the training set of the i^{th} class.

2. COMMON VECTOR APPROACH THROUGH LRC

To calculate the common vectors in CVA using LRC, it is better to start with the difference subspace of the i^{th} class. Let $\boldsymbol{a}_1^i, \boldsymbol{a}_2^i, \dots, \boldsymbol{a}_m^i,\ i = 1, \dots, C$ be the feature vectors of i^{th} class used in the training stage as defined before. The difference subspace of the i^{th} class is spanned by the difference vectors $\{\boldsymbol{a}_2^i - \boldsymbol{a}_1^i, \boldsymbol{a}_3^i - \boldsymbol{a}_1^i\ \dots, \boldsymbol{a}_m^i - \boldsymbol{a}_1^i\} = \{\boldsymbol{b}_1^i, \boldsymbol{b}_2^i, \dots, \boldsymbol{b}_{m-1}^i\}$. It is known that the subtrahend vector can be any of the feature vectors used in the training set [**5**].

The i^{th} distance metric of CVA can be written as

$$d_i = \left|(\boldsymbol{a}_x - \boldsymbol{a}_1^i) - \left(\widehat{\boldsymbol{a}_x - \boldsymbol{a}_1^i}\right)\right| = |\boldsymbol{b}_x - \widehat{\boldsymbol{b}}_x^i|. \tag{17}$$

$\widehat{\boldsymbol{b}}_x^i$ is the projection of $\boldsymbol{b}_x$ onto the difference subspace of the i^{th} class. If $\boldsymbol{P}_i$ is the projection matrix onto the difference subspace of the i^{th} class, then the above metric will be

$$d_i = |\boldsymbol{b}_x - \boldsymbol{P}_i \boldsymbol{b}_x|. \tag{18}$$

$\boldsymbol{P}_i$ can be determined in a similar way that we used in LRC estimation. Let $\boldsymbol{B}_i = [\boldsymbol{b}_1^i \vdots \boldsymbol{b}_2^i \vdots \cdots \vdots \boldsymbol{b}_{m-1}^i]$ be a matrix whose columns are the difference vectors of the i^{th} class.

$$\widehat{\boldsymbol{b}}_x^i = \boldsymbol{B}_i \boldsymbol{\beta}_{new,i} = \eta_1 \boldsymbol{b}_1^i + \eta_2 \boldsymbol{b}_2^i + \dots + \eta_{m-1} \boldsymbol{b}_{m-1}^i \tag{19}$$

or

$$\boldsymbol{b}_x = \boldsymbol{B}_i \boldsymbol{\beta}_{new,i} + \boldsymbol{\varepsilon}. \tag{20}$$

The sum of error squares is

$$S = \boldsymbol{\varepsilon}^T \boldsymbol{\varepsilon} = (\boldsymbol{b}_x - \boldsymbol{B}_i \boldsymbol{\beta}_{new,i})^T (\boldsymbol{b}_x - \boldsymbol{B}_i \boldsymbol{\beta}_{new,i}). \tag{21}$$

After the minimization process the estimated vector parameters are obtained as

$$\boldsymbol{\beta}_{new,i} = \left(\boldsymbol{B}_i^T\boldsymbol{B}_i\right)^{-1}\boldsymbol{B}_i^T\boldsymbol{b}_x. \tag{22}$$

If we combine Eq.(17), Eq.(19) and Eq.(22), we will get

$$\boldsymbol{P}_i = \boldsymbol{B}_i\left(\boldsymbol{B}_i^T\boldsymbol{B}_i\right)^{-1}\boldsymbol{B}_i^T. \tag{23}$$

Let $\boldsymbol{P}_i^{\perp}$ denotes the projection matrix onto the indifference subspace. Then Eq.(18) becomes

$$\begin{aligned} d_i &= |\boldsymbol{b}_x - \boldsymbol{P}_i\boldsymbol{b}_x| = \left|\boldsymbol{P}_i^{\perp}\boldsymbol{b}_x\right| = \left|\boldsymbol{P}_i^{\perp}(\boldsymbol{a}_x - \boldsymbol{a}_1)\right| = \left|\boldsymbol{P}_i^{\perp}\boldsymbol{a}_x - \boldsymbol{P}_i^{\perp}\boldsymbol{a}_1\right| \\ &= \left|\boldsymbol{P}_i^{\perp}\boldsymbol{a}_x - \boldsymbol{a}_{com}^i\right| \end{aligned} \tag{24}$$

where $\boldsymbol{a}_{com}^i$ is the common vector of the i^{th} class. Therefore the metric in Eq.(18) gives the distance between the projection of the unknown feature vector $\boldsymbol{a}_x$ onto the indifference subspace of the i^{th} class and its common vector. Whenever $\boldsymbol{a}_x$ is one of the feature vectors of the i^{th} class in the training set, then $\boldsymbol{P}_i^{\perp}\boldsymbol{a}_x$ is always equal to the common vector $\boldsymbol{a}_{com}^i$ [5] and d_i is zero.

Numerical Example

Let $\boldsymbol{a}_1 = [1 \quad 2 \quad -1 \quad 0]^T$, $\boldsymbol{a}_2 = [-1 \quad 1 \quad 4 \quad 1]^T$, $\boldsymbol{a}_3 = [3 \quad 0 \quad -1 \quad 4]^T$, be the feature vectors of a class.

(i) To use the Gram-Schmidt orthogonalization procedure, let the difference vectors be

$$\begin{aligned} \boldsymbol{b}_1 &= \boldsymbol{a}_2 - \boldsymbol{a}_1 = [-2 \quad -1 \quad 5 \quad 1\,]^T, \\ \boldsymbol{b}_2 &= \boldsymbol{a}_3 - \boldsymbol{a}_1 = [2 \quad -2 \quad 0 \quad 4\,]^T. \end{aligned}$$

If the difference vectors are orthonormalized using the Gram-Schmidt orthogonalization procedure, then the orthonormal vectors become,

$$\begin{aligned} \boldsymbol{z}_1 &= [-0.3592 \quad -0.1796 \quad 0.8980 \quad 0.1796]^T \\ \boldsymbol{z}_2 &= [0.4358 \quad -0.3961 \quad -0.0660 \quad 0.8055]^T \end{aligned}$$

The common vector of the class is calculated as below:

$$\boldsymbol{a}_{com} = \boldsymbol{a}_i - (\boldsymbol{z}_1^T\boldsymbol{a}_i)\boldsymbol{z}_1 - (\boldsymbol{z}_2^T\boldsymbol{a}_i)\boldsymbol{z}_2 = [0.5459 \quad 1.5946 \quad 0.4324 \quad 0.5243]^T,\ i = 1,2,3$$

Here the projection matrix onto the difference subspace is the following

$$\boldsymbol{P} = \boldsymbol{z}_1\boldsymbol{z}_1^T + \boldsymbol{z}_2\boldsymbol{z}_2^T = \begin{bmatrix} 0.3189 & -0.1081 & -0.3514 & 0.2865 \\ -0.1081 & 0.1892 & -0.1351 & -0.3514 \\ -0.3514 & -0.1351 & 0.8108 & 0.1081 \\ 0.2865 & -0.3514 & 0.1081 & 0.6811 \end{bmatrix}.$$

The projection matrix onto the indifference subspace is given below:

$$\boldsymbol{P}^{\perp} = \boldsymbol{I} - \boldsymbol{P} = \begin{bmatrix} 0.6811 & 0.1081 & 0.3514 & -0.2865 \\ 0.1081 & 0.8108 & 0.1351 & 0.3514 \\ 0.3514 & 0.1351 & 0.1892 & -0.1081 \\ -0.2865 & 0.3514 & -0.1081 & 0.3189 \end{bmatrix}$$

(*ii*) Secondly, common vector of the class can be calculated using covariance matrix of that class which is called as the within-class covariance matrix. The within-class covariance matrix of the class is

$$\boldsymbol{\Phi} = \begin{bmatrix} 4 & -1 & -5 & 3 \\ -1 & 1 & 0 & -2 \\ -5 & 0 & 8.33 & -1.67 \\ 3 & -2 & -1.67 & 4.33 \end{bmatrix}.$$

By applying the eigenvalue-eigenvector decomposition to**,** we obtain the following pairs:

$$\begin{aligned} \boldsymbol{u}_1 &= [-0.8177 \quad -0.2500 \quad -0.4339 \quad 0.2839]^T, & \lambda_1 &= 0 \\ \boldsymbol{u}_2 &= [0.1113 \quad -0.8651 \quad -0.0308 \quad -0.4882]^T, & \lambda_2 &= 0 \\ \boldsymbol{u}_3 &= [0.1151 \quad -0.4214 \quad 0.5105 \quad 0.7407]^T, & \lambda_3 &= 4.7884 \\ \boldsymbol{u}_4 &= [0.5529 \quad 0.1078 \quad -0.7417 \quad 0.3640]^T, & \lambda_4 &= 12.8782 \end{aligned}$$

Here the projection matrix of the indifference subspace is

$$\boldsymbol{P}^{\perp} = \boldsymbol{u}_1^T\boldsymbol{u}_1 + \boldsymbol{u}_2^T\boldsymbol{u}_2 = \begin{bmatrix} 0.6811 & 0.1081 & 0.3514 & -0.2865 \\ 0.1081 & 0.8108 & 0.1351 & 0.3514 \\ 0.3514 & 0.1351 & 0.1892 & -0.1081 \\ -0.2865 & 0.3514 & -0.1081 & 0.3189 \end{bmatrix}.$$

The common vector has the same values as before

$$\boldsymbol{a}_{com} = \boldsymbol{P}^{\perp}\boldsymbol{a}_i = [0.5459 \quad 1.5946 \quad 0.4324 \quad 0.5243]^T, i = 1,2,3.$$

(*iii*) If we form the difference matrix as $\boldsymbol{B} = [\boldsymbol{b}_1 \vdots \boldsymbol{b}_2]$ and substitute it Eq.(23) in LRC method, we obtain the projection matrix onto the difference subspace of the class:

$$\boldsymbol{P} = \begin{bmatrix} 0.3189 & -0.1081 & -0.3514 & 0.2865 \\ -0.1081 & 0.1892 & -0.1351 & -0.3514 \\ -0.3514 & -0.1351 & 0.8108 & 0.1081 \\ 0.2865 & -0.3514 & 0.1081 & 0.6811 \end{bmatrix}$$

Then $\boldsymbol{P}^{\perp}$ is calculated using the equality $\boldsymbol{P} + \boldsymbol{P}^{\perp} = \boldsymbol{I}$. The common vector of the class is calculated using $\boldsymbol{P}^{\perp}$: $\boldsymbol{a}_{com} = \boldsymbol{P}^{\perp}\boldsymbol{a}_i{}^T = [0.5459 \quad 1.5946 \quad 0.4324 \quad 0.5243]$, $i = 1,2,3$.

In the example, we see that the projection matrices $\boldsymbol{P}$ and $\boldsymbol{P}^{\perp}$, and the common vectors obtained by using three algorithms are exactly the same.

3. DISCRIMINATIVE COMMON VECTOR APPROACH THROUGH LRC

Let C is the number of classes and m is the number of feature vectors from each class as before. Then the whole set of difference vectors is obtained from all the difference vectors belonging to all the classes as below

$$\boldsymbol{B} = \{\boldsymbol{b}_1^1, \boldsymbol{b}_2^1, \dots, \boldsymbol{b}_{m-1}^1, \dots, \boldsymbol{b}_1^2, \boldsymbol{b}_2^2, \dots, \boldsymbol{b}_{m-1}^2, \dots, \boldsymbol{b}_1^C, \boldsymbol{b}_2^C, \dots, \boldsymbol{b}_{m-1}^C\} \tag{25}$$

where $\boldsymbol{b}_j^i = \boldsymbol{a}_j^i - \boldsymbol{a}_1^i, i = 1, \dots, C,\ j = 2, \dots, m$. The difference vectors in the set $\boldsymbol{B}$ span the difference subspace in DCVA. Let $\boldsymbol{a}_x$ be the unknown feature vector which is going to be classified. Similar with CVA, the distance metric can be written as

$$d_i = \left|(\boldsymbol{a}_x - \boldsymbol{a}_1^i) - (\widehat{\boldsymbol{a}_x - \boldsymbol{a}_1^i})\right| = \left|\boldsymbol{b}_x - \widehat{\boldsymbol{b}}_x^i\right|. \tag{26}$$

The projection of $\boldsymbol{b}_x$ onto the difference subspace can be written as $\widehat{\boldsymbol{b}}_x^i = \boldsymbol{B}\boldsymbol{\beta}_i$. Here $\boldsymbol{\beta}_i$ can be estimated using least squares estimation similar with LRC method,

$$\boldsymbol{\beta}_i = (\boldsymbol{B}^T\boldsymbol{B})^{-1}\boldsymbol{B}\boldsymbol{b}_x. \tag{27}$$

Then

$$\widehat{\boldsymbol{b}}_x^i = \boldsymbol{P}\boldsymbol{b}_x \tag{28}$$

where $\boldsymbol{P} = \boldsymbol{B}(\boldsymbol{B}^T\boldsymbol{B})^{-1}\boldsymbol{B}$. Its complement $\boldsymbol{P}^{\perp}$ can becalculated using the equality $\boldsymbol{P} + \boldsymbol{P}^{\perp} = \boldsymbol{I}$.

Finally the distance in Eq.(26) becomes

$$\begin{aligned} d_i &= \left|\boldsymbol{b}_x - \widehat{\boldsymbol{b}}_x^i\right| = \left|(\boldsymbol{a}_x - \boldsymbol{a}_1^i) - \boldsymbol{B}(\boldsymbol{B}^T\boldsymbol{B})^{-1}\boldsymbol{B}(\boldsymbol{a}_x - \boldsymbol{a}_1^i)\right| \\ &= \left|\boldsymbol{a}_x - \boldsymbol{B}(\boldsymbol{B}^T\boldsymbol{B})^{-1}\boldsymbol{B}\boldsymbol{a}_x - \left(\boldsymbol{a}_1^i - \boldsymbol{B}(\boldsymbol{B}^T\boldsymbol{B})^{-1}\boldsymbol{B}\boldsymbol{a}_1^i\right)\right| \\ &= \left|\boldsymbol{P}^{\perp}\boldsymbol{a}_x - \boldsymbol{a}_{com}^i\right|. \end{aligned} \tag{29}$$

In the first step, the common vectors of the classes should be calculated using the following formula

$$\begin{aligned} \boldsymbol{a}_{com}^i &= \boldsymbol{a}_j^i - \boldsymbol{B}(\boldsymbol{B}^T\boldsymbol{B})^{-1}\boldsymbol{B}\boldsymbol{a}_j^i \; i = 1, \dots, C, \qquad j = 1, \dots, m \\ &= \boldsymbol{a}_j^i - \boldsymbol{P}\boldsymbol{a}_j^i \; i = 1, \dots, C, \qquad j = 1, \dots, m \\ &= \boldsymbol{P}^{\perp}\boldsymbol{a}_j^i \; i = 1, \dots, C, \qquad j = 1, \dots, m \end{aligned} \tag{30}$$

The rest of the derivations of the discriminative common vectors are the same as in Section 1.2.

Numerical Example

In this numerical example we have two classes and each class has two feature vectors in four dimensional space. Let $\boldsymbol{a}_1^1 = [1 \quad 2 \quad -1 \quad 0]^T$, $\boldsymbol{a}_2^1 = [-1 \quad 1 \quad 4 \quad 1]^T$ be the feature vectors of class C_1 and $\boldsymbol{a}_1^2 = [3 \quad -4 \quad 0 \quad 2]^T$, $\boldsymbol{a}_2^2 = [2 \quad -3 \quad 2 \quad -1]^T$ be the feature vectors of class C_2.

(*i*) The difference vectors:$\boldsymbol{b}_1 = \boldsymbol{a}_2^1 - \boldsymbol{a}_1^1 = [-2 \quad -1 \quad 5 \quad 1]^T$, $\boldsymbol{b}_2 = \boldsymbol{a}_2^2 - \boldsymbol{a}_1^2 = [-1 \quad 1 \quad 2 \quad -3]^T$. After the Gram-Schmidt orthogonalization process, the normalized difference vectors will become

$$\begin{aligned} \boldsymbol{z}_1 &= [-0.3592 \quad -0.1796 \quad 0.8980 \quad 0.1796]^T, \\ \boldsymbol{z}_2 &= [-0.1345 \quad 0.3498 \quad 0.1973 \quad -0.9059]^T. \end{aligned}$$

The common vectors of the classes become

$$\begin{aligned} \boldsymbol{a}_{com}^1 &= \boldsymbol{a}_i^1 - (\boldsymbol{z}_1^T\boldsymbol{a}_i^1)\boldsymbol{z}_1 - (\boldsymbol{z}_2^T\boldsymbol{a}_i^1)\boldsymbol{z}_2 = [0.4688 \quad 1.5810 \quad 0.3791 \quad 0.6234]^T, i = 1,2 \\ \boldsymbol{a}_{com}^2 &= \boldsymbol{a}_i^2 - (\boldsymbol{z}_1^T\boldsymbol{a}_i^2)\boldsymbol{z}_1 - (\boldsymbol{z}_2^T\boldsymbol{a}_i^2)\boldsymbol{z}_2 = [2.5137 \quad -2.7357 \quad 0.7132 \quad -1.2743]^T, i = 1,2. \end{aligned}$$

Here the projection matrices of the difference and the indifference subspaces are

$$\boldsymbol{P} = \begin{bmatrix} 0.1471 & 0.0175 & -0.3491 & 0.0574 \\ 0.0175 & 0.1546 & -0.0923 & -0.3491 \\ -0.3491 & -0.0923 & 0.8454 & -0.0175 \\ 0.0574 & -0.3491 & -0.0175 & 0.8529 \end{bmatrix}$$

and

$$\boldsymbol{P}^{\perp} = \begin{bmatrix} 0.8529 & -0.0175 & 0.3491 & -0.0574 \\ -0.0175 & 0.8454 & 0.0923 & 0.3491 \\ 0.3491 & 0.0923 & 0.1546 & 0.0175 \\ -0.0574 & 0.3491 & 0.0175 & 0.1471 \end{bmatrix}$$

respectively.

(*ii*) The common vectors of the classes also can be found by using the total within-class covariance matrix. Let $\boldsymbol{\Phi}_1$ and $\boldsymbol{\Phi}_2$ be the covariance matrices of C_1 and C_2 respectively. The total within-class covariance matrix can be calculated as

$$\boldsymbol{\Phi} = \boldsymbol{\Phi}_1 + \boldsymbol{\Phi}_2 = \begin{bmatrix} 2.5 & 0.5 & -6 & 0.5 \\ 0.5 & 1 & -1.5 & -2 \\ -6 & -1.5 & 14.5 & -0.5 \\ 0.5 & -2 & -0.5 & 5 \end{bmatrix}.$$

By applying eigenvalue-eigenvector decomposition to $\boldsymbol{\Phi}$ the following pairs will be obtained:

$$\begin{aligned} \boldsymbol{u}_1 &= [-0.5632 \quad -0.7170 \quad -0.3163 \quad -0.2621]^T, & \lambda_1 &= 0, \\ \boldsymbol{u}_2 &= [-0.7319 \quad 0.5756 \quad -0.2336 \quad 0.2801]^T, & \lambda_2 &= 0, \\ \boldsymbol{u}_3 &= [-0.0464 \quad 0.3822 \quad -0.0192 \quad -0.9227]^T, & \lambda_3 &= 5.8431, \\ \boldsymbol{u}_4 &= [-0.3808 \quad -0.0924 \quad 0.9192 \quad -0.0383]^T, & \lambda_4 &= 17.1569. \end{aligned}$$

The projection matrix onto the indifference subspace is formed using the eigenvectors corresponding to the zero eigenvalues.

$$\boldsymbol{P}^{\perp} = \begin{bmatrix} 0.8529 & -0.0175 & 0.3491 & -0.0574 \\ -0.0175 & 0.8454 & 0.0923 & 0.3491 \\ 0.3491 & 0.0923 & 0.1546 & 0.0175 \\ -0.0574 & 0.3491 & 0.0175 & 0.1471 \end{bmatrix}$$

Then the common vectors are,

$$\begin{aligned} \boldsymbol{a}_{com}^1 &= \boldsymbol{P}^{\perp}\boldsymbol{a}_i^1 = [0.4688 \quad 1.5810 \quad 0.3791 \quad 0.6234]^T, & i &= 1,2 \\ \boldsymbol{a}_{com}^2 &= \boldsymbol{P}^{\perp}\boldsymbol{a}_i^2 = [2.5137 \quad -2.7357 \quad 0.7132 \quad -1.2743]^T, & i &= 1,2 \end{aligned}$$

(*iii*) In the proposed method, the difference matrix $\boldsymbol{B} = [\boldsymbol{b}_1 \vdots \boldsymbol{b}_2]$ is formed as it is substituted in Eq.(23). Finally the projection matrix onto the difference subspace becomes

$$\boldsymbol{P} = \begin{bmatrix} 0.1471 & 0.0175 & -0.3491 & 0.0574 \\ 0.0175 & 0.1546 & -0.0923 & -0.3491 \\ -0.3491 & -0.0923 & 0.8454 & -0.0175 \\ 0.0574 & -0.3491 & -0.0175 & 0.8529 \end{bmatrix}.$$

The common vectors are

$$\begin{aligned} \boldsymbol{a}_{com}^1 &= \boldsymbol{a}_i^1 - \boldsymbol{P}\boldsymbol{a}_i^1 = [0.4688 \quad 1.5810 \quad 0.3791 \quad 0.6234]^T, & i &= 1,2 \\ \boldsymbol{a}_{com}^2 &= \boldsymbol{a}_i^2 - \boldsymbol{P}\boldsymbol{a}_i^2 = [2.5137 \quad -2.7357 \quad 0.7132 \quad -1.2743]^T, & i &= 1,2 \end{aligned}$$

Thus, the same common vectors are obtained using the above three algorithms. The next step is to obtain the discriminative common vectors. The difference of the common vectors are obtained as $\boldsymbol{b}_{com}^1 = \boldsymbol{a}_{com}^2 - \boldsymbol{a}_{com}^1 = [2.0449 \quad -4.3167 \quad 0.3342 \quad -1.8978]$.

Then the projection matrix $\boldsymbol{U}$ onto the difference subspace of the common vectors $\boldsymbol{U} = [0.3970 \quad -0.8381 \quad 0.0649 \quad -0.3685]^T$ which is equal to the normalized difference vector between the commons, $\boldsymbol{U} = \frac{b_{com}^1}{\|b_{com}^1\|}$. Then the discriminative common vectors of the classes are one dimensional and $\boldsymbol{\Omega}_1 = \boldsymbol{U}^T \boldsymbol{a}_i^1 = -6.9227, i = 1,2$ and $\boldsymbol{\Omega}_2 = \boldsymbol{U}^T \boldsymbol{a}_i^2 = 19.6060\ , i = 1,2$.

Figure 2. Preprocessed images of two subjects from AR face database

4. EXPERIMENTAL WORK

In addition to the numerical examples, a simple face recognition case study is performed to see the equivalence of the different realizations of CVs and the equivalence of the different implementations of DCVs using AR face database [15]. As an additional analysis, we compare recognition performances of CVA, DCVA, and LRC in this face recognition task. AR face database contains 26 images taken in two sessions including different facial expressions (neutral, angry, scream, smile), illumination conditions (left light on, right light on, all sides light on), and occlusions (sun glass, scarf) from 126 individuals. The original size of the images is 768×576. Images are converted to grayscale, cropped and normalized such that eyes of each individual are at the same coordinate and resized to 120×90. In the experiments we used 100 subjects (64 male and 36 female). Images of two individuals obtained after the preprocessing operations from AR database are shown in Figure 2.We randomly choose 7 images from each subject for training set. Thus 700 images are used for training purposes. The rest of the images are used for testing. This process is repeated 10 times and the recognition rates are obtained by averaging each run. The recognition rates using three different implementations of CVs (DCVs) are the same as expected. The recognition performances of LRC, CVA, and DCVA with the standard deviations are shown in Table 1. As seen from Table 1, DCVA outperforms LRC and CVA as reported in [3,16]. Also DCVA is more robust classification method, since the standard deviation of DCVA is smaller than the others.

Table 1. Recognition performances of LRC, CVA, and DCVA in AR face database

LRC	CVA	DCVA
90.03 ± 1.10	88.21 ± 1.65	95.97 ± 0.55

We use the above scenario to compare the training and the testing times of the different implementations of common vectors. In Table 2, *train* and *test* represent the training time of the system and the testing time of a query image respectively. The results are given in milliseconds. The training time of LRC-based algorithm is the best among three. But LRC-based algorithm gives the worst results in terms of testing time because of the matrix inversion and additional matrix multiplications in Eq.(23).

Table 2. Training and testing times of three implementation algorithms for common vectors

	Gram-Schmidt-based	Covariance matrix-based	LRC-based
Train (ms)	679.2	708.1	514.4
Test (ms)	14.35	14.64	34.23

5. CONCLUSION

There are two different methods to obtain common vectors and discriminative common vectors, i.e. Gram-Schmidt orthogonalization or covariance analysis. In this paper, we proposed third algorithm to obtain CVs and DCVs by using the linear regression idea. We gave two synthetic examples to clarify the proposed method. In these experiments, the common vectors calculated from three different algorithms are exactly the same, so are the recognition rates as expected. Also, we present a face recognition experiment in AR face database. In the experiment, it is seen that DCVA has better classification performance than CVA and LRC as reported in the literature [16]. We show that the proposed implementation algorithm is the faster than the other two in the training phase. But it has higher computational complexity when compared with the conventional implementation algorithms in the testing stage.

Each method has its own advantages as well as disadvantages. For example for applying the Gram-Schmidt orthogonalization process a recursive relation is established which may need an extra longer time, compared with other methods, for arithmetical operations. Also a residual error may be building up at each recursion step. Covariance matrix based calculations can be used to overcome these difficulties. The calculations of the covariance matrices are not that difficult in the second method, eigenvalue-eigenvector decomposition of these matrices can be troublesome due to high dimensionalities. There are tricks to overcome this difficulty [**3**]. Finally one needs to take inverse of high dimensional matrices sometimes to calculate the projection matrices in LRC unless orthonormalized column vectors are calculated beforehand.

REFERENCES

[1] Shakhnarovich G, Moghaddam B. Face recognition in subspaces In: Li SZ, Jain AK, editors. Handbook of Face Recognition. Springer-Verlag; 2004. pp. 141-168.

[2] Gulmezoglu MB, Dzhafarov V, Barkana A. The common vector approach and its relation to principal component analysis. IEEE T Speech and Audi P 2001; 9(6): 655-662.

[3] Cevikalp H, Neamtu M, Wilkes M, Barkana A. Discriminative common vectors for face recognition. IEEE T Pattern Anal 2005; 27(1): 4-13.

[4] Naseem I, Togneri R, Bennamoun M. Linear regression for face recognition. IEEE T Pattern Anal 2010; 32(11): 2106-2112.

[5] Koc M, Barkana A, Gerek ON. A fast method for the implementation of common vector approach. Inform Sciences 2010; 180(11): 4084-4098.

[6] Gulmezoglu MB, Keskin M, Dzhafarov V, Barkana A. A novel approach to isolated word recognition. IEEE T Speech and Audi P 1999; 7(6): 620-628.

[7] Lu GF, Zou J, Wang Y. Incremental learning of discriminant common vectors for feature extraction. Appl Math Comput 2012; 218(22): 11269-11278.

[8] Koc M, Barkana A. An implementation of discriminative common vector approach using matrices. In: The Seventh International Multi-Conference on Computing in the Global Information Technology(ICCGI2012); 24-29 June 2012; Venice, Italy. pp. 260-263.

[9] Cevikalp H, Neamtu M, Barkana A. The kernel common vector method: a novel nonlinear subspace classifier for pattern recognition. IEEE T Syst Man Cy B 2007; 37(4): 937-951.

[10] Diaz-Chito K, Ferri FJ, Diaz-Villanueva W. Image recognition through incremental discriminative common vectors. Advenced Concepts for Intelligent Vision Systems 2010; 6475: 304-311.

[11] Huang SM, Yang JF. Linear discriminant regression classification for face recognition. IEEE Signal Proc Let 2013; 20(1): 91-94.

[12] Huang SM, Yang JF. Improved principal component regression for face recognition under illumination variations. IEEE Signal Proc Let 2012; 19(4): 179-182.

[13] Naseem I, Togneri R, Bennamoun M. Robust regression for face recognition. Pattern Recogn 2012; 45(1): 104-118.

[14] Koc M, Barkana A. Application of linear regression classification to low dimensional datasets. Neurocomputing 2014; 131: 331-335.

[15] Martinez A, Benavente Y. The AR face database, CVC Technical Report 24, 1994.

[16] Koc M, Barkana A. Discriminative common vector approach based feature selection in face recognition, Computers & Electrical Engineering, 40(8):37-50, 2014.