DISCRIMINATIVE COMMON VECTORS BASED ON THE GRAM-SCHMIDT REORTHOGONALIZATION FOR THE SMALL SAMPLE SIZE PROBLEM

Ying Wen[1]*, Lianghua He*[2]*, Yue Lu*[1]

1. Department of Computer Science and Technology, East China Normal University, Shanghai 200062, China

2. Key Laboratory of Embedded System and Service Computing, Ministry of Education, Tongji University, Shanghai 200092, China

ABSTRACT

The discriminative common vectors (DCV) algorithm shows better face recognition effects than some commonly used linear discriminant algorithms, which uses the subspace methods and the Gram-Schmidt orthogonalization (GSO) procedure to obtain the DCV. However, the Gram-Schmidt technique may produce a set of vectors which is far from orthogonal so that sometimes the orthogonality may be lost completely. Hence, the effectiveness of the DCV is also decreased. In this paper, we proposed an improved DCV method based on the GSO. For obtaining an accurate projection onto the corresponding space, the orthogonal basis problem is usually solved with the Gram-Schmidt process with reorthogonalization. Thus, the effectiveness of the DCV can be improved and the experimental results show that the proposed method is better for the small sample size problem as compared to the DCV.

Index Terms**—** Face recognition, Discriminative common vector, Gram-Schmidt orthogonalization

1. INTRODUCTION

Small sample size (SSS) problem often presents in the pattern recognition filed, especially in image recognition problem. In general, image capture for enough samples is often difficulty in practical environment, such as face image, finger image and signature and so on. However, even if we obtain these images, the sample size is often small and only 2 or 3 samples are in each class [1]. Most of the statistical methods suffer from this problem. In order to do research on the SSS problem, we often use affine transformation to expand the samples, which includes translation transform, scaling transform, shear transform and rotation transform. Then, we do the recognition on the expended samples.

Recently, the common vector was proposed and originally introduced for isolated word recognition problems in which the number of samples in each class was less than or equal to the dimensionality of the sample space [2]. The common vector presents the common properties of a training set [3]. Inspired by this idea, Cevikalp et al. [4] proposed the discriminative common vectors (DCV) for face recognition, which uses the subspace methods and the Gram-Schmidt orthogonalization (GSO) procedure to obtain the DCV, and shows better face recognition effects than some commonly used linear discriminant algorithms. However, it was confirmed by many numerical experiments, that the classical GSO may produce a set of vectors which is far from orthogonal and sometimes the orthogonality can be lost completely [5], [6]. Nevertheless, despite its weakness, this technique is frequently implemented due to its simplicity and potential parallelism. As we know, the orthogonality of the computed vectors is essential for obtaining an accurate projection onto the corresponding space [7], [8]. If the projection vectors in the difference subspace are not orthogonal, the DCV obtained by the common vectors GSO procedure are also not optimal. Yet, to solve the SSS problem, the expanded samples are obtained by the affine transformation in our paper so that the linear correlation between the samples is unavoidable. To obtain the orthogonal vectors in these expanded samples, researchers proposed the modified Gram-Schmidt (MGS) [9], modified Gram-Schmidt with pivoting (MGS-Pivot) [10] and iterative Gram-Schmidt orthogonalization [11].

In this paper, we proposed an improved the DCV using an iterative GSO to obtain the orthogonal projective vectors in the difference subspace so that the DCV's effectiveness can be improved. Experiments show that the performance of the proposed algorithm is the same as that of the DCV when the columns of the subspace have linear independency, while the performance of the proposed algorithm is better than that of the DCV when the columns of the subspace have linear correlation. This paper is organized as follows. In Section 2 an improved DCV based on the Gram-Schmidt reorthogonalization is described. Experimental results on the ORL and AR face database are shown in Section 3. Finally, the conclusion is given in Section 4.

2. DISCRIMINATIVE COMMON VECTORS BASED ON THE GRAM-SCHMIDT REORTHOGONALIZATION

2.1 Discriminative common vectors

The DCV method is based on a variation of Fisher's Linear Discriminant Analysis for the small sample size cases. Two different algorithms were given to extract the

DCV representing each person in the training set of the face database. One algorithm uses the within-class scatter matrix of the samples in the training set, while the other uses the subspace methods and the GSO to obtain the DCV. In this paper, we focus on the DCV based on the difference subspace and the GSO procedure.

Let a training set be composed of *C* class, where each class contains *N* samples, and let x_m^i be a d-dimensional column vector which denotes the *mth* sample from the *ith* class. There will be a total of *M=NC* samples in the training set (suppose that *d>M-C*). In the DCV, the subspace methods can be applied to obtain the common vectors x_{com}^i for each class *i*. We chose any one of the image vectors from the *ith* class as the subtrahend vector, and then obtained the difference vectors b_k^i of the so-called difference subspace of the *ith* class. Thus, assuming that the first sample of each class is taken as the subtrahend vector, we have

$$b_k^i = x_{k+1}^i - x_1^i, k = 1, ..., N-1 \quad (1)$$

The difference subspace B_i of the *ith* class is defined as $B_i = span\{b_1^i, ... b_{N-1}^i\}$. These subspaces can be summed up to form the complete difference subspace as defined below:

$$B = B_1 + ... + B_C = span\{b_1^1, ..., b_{N_1-1}^1, b_1^2, ..., b_{N_C-1}^C\} \quad (2)$$

The GSO is used to obtain orthonormal basis vectors to locate the projection matrix onto *B*. However, we notice that the difference subspace *B,* when its *det(B'B)* is close to zero, the orthogonality of the computed vectors may be significantly lost in the GSO algorithm [5], [6]. If the projection vectors in the difference subspace are not orthogonal, the common vectors may be deviated. The DCV obtained by the common vectors Gram-Schmidt procedure, which are important for the face recognition results, are sub-optimal. In the following subsection, we analyzed the loss of orthogonality in the GSO.

2.2 Gram-Schmidt reorthogonalization (GSR)

In this section, we present a brief description of the GSO process. Assume we have a set $A = \{a_i \mid i = 1, 2, ..., n\}$ of *m*-vectors and wish to obtain an equivalent orthonormal set $Q = \{q_i \mid i = 1, 2, ..., n\}$ of *m*-vectors. (Table 1)

Table 1 The GSO and GSR.

GSO	GSR

for j=1->n do
$w = (\mathrm{I}_m - Q_{j-1}Q_{j-1}^T)a_j$
$q_j = w / \lVert w \rVert$
end for
	for j=1->n do
	$w = a_j$
	for *r*=1->2 do
	$w = (I_m - Q_{j-1}Q_{j-1}^T)w$
	end for
	$q_j = w / \lVert w \rVert$
	end for

In general, the loss of orthogonality (LOO) of vectors *Q* computed in the GSO can be bounded as [5]:

$$\| I - Q'Q \| \le \gamma \varepsilon \kappa(A) \quad (3)$$

Where ε is a constant that depends only on the details of the computer's arithmetic, then ε is the machine precision (or unit round-off) and $\kappa(A)$ is the condition number of matrix *A*. The bound on the LOO of computed vectors is proportional to $\kappa(A)$.

If the determinant value of *A'A*, *det(A'A)* is close to zero, the inverse of the *A'A*, *det(A'A)* will be extremely inflated, and in consequence the solution analysis will have poor robustness and low precision. In this case, matrix *A* can be regarded as ill-conditioned. The ill-conditioned matrix results from linear correlation among columns in the *A* matrix. For ill-conditioned matrices, the computed vectors can be very far from orthogonal.

Let us present an example to show the LOO of the Gram-Schmidt algorithm. A Hilbert matrix is a typical ill-conditioned matrix and is defined as:

$$A = (a_{ij}) = 1/(i+j-1); i, j = 1, 2, ..., n \quad (4)$$

Let A be the 12 x 12 Hilbert matrix:

$$A = hilb(12) = \begin{bmatrix} 1 & 1/2 & ... & 1/12 \\ 1/2 & 1/3 & ... & 1/13 \\ \vdots & \vdots & \ddots & \vdots \\ 1/12 & 1/13 & ... & 1/23 \end{bmatrix} \quad (5)$$

We have *k(A)*=3.8e+16. With the GSO, we get $\| I - Q'Q \|$=8.9, so *Q* is far from being orthogonal. In general, computation in floating-point arithmetic implies round-off errors. These errors in the GSO may lead to a severe LOO in the *Q*-factor. In theory, if all vectors are orthogonal, the value of $\| I - Q'Q \|$ is close to zero. The projection vectors of the Hilbert matrix obtained by the classical GSO are not completely orthogonal.

However, it is important to compute the vectors *Q* so that their orthogonality is close to the machine precision level. The orthogonality of the computed vectors is essential for obtaining an accurate projection onto the corresponding space [7], [8]. In general, the orthogonal basis problem is usually solved with the GSR. Theoretically, reorthogonalization could be applied several times, but in a practical application, one reorthogonalization of a current vector against the previously computed set is performed exactly twice [6]. In the right column of Table 1, the GSR is presented. The difference between the GSO and GSR lies in the computation times of the orthogonal vectors.

With the Gram-Schmidt algorithm iterated twice on the above Hibert matrix, we get $\| I - Q'Q \|$=2.9e-16 which is close to zero and far less than 8.9 obtained by the Gram-Schmidt algorithm. Hence, the orthogonalizaion of the Gram-Schmidt algorithm could be solved by the Gram-Schmidt iteration algorithm.

We then analyzed the selective reothogonalization criterion. The Gram–Schmidt algorithm has computed matrices Q and R that satisfy a bound of the form:

$$\|A - QR\| \sim \gamma\varepsilon \|A\| \quad (6)$$

where ε denotes the machine precision (or unit round-off) in our computations and ε is a constant that depends on the dimension of Q and the details of the arithmetic. This property is always maintained in the Gram-Schmidt algorithm.

At step j of the Gram-Schmidt algorithm, we compute w such that $w = (I_m - Q_{j-1}Q_{j-1}^T)a_j$. In floating-point arithmetic, we compute $\tilde{w}$ such that

$$\tilde{w} = (I_m - Q_{j-1}Q_{j-1}^T)a_j + e \quad (7)$$

where $\|e\| \sim \gamma\varepsilon \|a_j\|$. This gives

$$Q_{j-1}^T \tilde{w} = Q_{j-1}^T e \quad (8)$$

Computing $\tilde{q}_j = \tilde{w} / \|\tilde{w}\|$ leads to

$$\|Q_{j-1}^T \tilde{q}_j\| \sim \gamma\varepsilon \left(\|a_j\| / \|w\|\right) \quad (9)$$

At step j, the quantity $\|a_j\| / \|w\|$ clearly controls the LOO $\tilde{q}_j$ against Q_{j-1}. When $\|a_j\| / \|w\| < K$, then the reorthogonalization procedure ends. In general, K is defined as $1 \le K \le \kappa(A)$ [12].

In this paper, we proposed the DCV using the GSR for face recognition. The procedure is as follows:

Step 1: Calculate the range space of the within-class matrix, which is identical to the range space of the difference subspace B. Here, B is defined as

$$B = [b_1^1, ..., b_{N_1-1}^1, b_1^2, ..., b_{N_C-1}^C] \quad (10)$$

where $b_j^i = x_j^i - x_{N_i}^i, i = 1, ..., C, j = 1, ..., N_i - 1$ is the *jth* difference vector of the *ith* class. For B, apply the GSR procedure and set a reothogonalization criterion, and then get $B = UV$. Then, U is an orthonormal matrix whose column vectors span the range space of the within-class matrix.

Step 2: Choose any sample from each class (typically, the last sample of the *ith* class $x_{N_i}^i$) and project it to the null space of the within-class matrix through the following equation

$$x_{com}^i = x_j^i - UU' x_j^i = x_{N_i}^i - UU' x_{N_i}^i \quad (11)$$

where x_{com}^i is a common vector of the *ith* class and is independent of index j.

Step 3: From the matrix B_{com} where

$$B_{com} = [b_{com}^1, b_{com}^2, ..., b_{com}^{C-1}]$$
$$b_{com}^i = x_{com}^i - x_{com}^C, i = 1, ..., C-1 \quad (12)$$

Finally, we apply the GSO to B_{com} and then get the projection matrix.

3. EXPERIMENTS

The ORL and AR face databases are used to test the proposed method.

3.1 Experiments with the ORL face database

In this group of experiments, the Olivetti-Oracle Research Lab (ORL) face database is adopted and used to test the performance of face recognition algorithms under the condition of minor variation of scaling and rotation (as shown in Fig.1). The ORL face database contains 400 frontal faces sized 112 x 92: 10 tightly, cropped images of 40 individuals with variation in pose, illumination, facial expression (open/closed eyes, smiling/not smiling) and facial details (glasses/no glasses) .

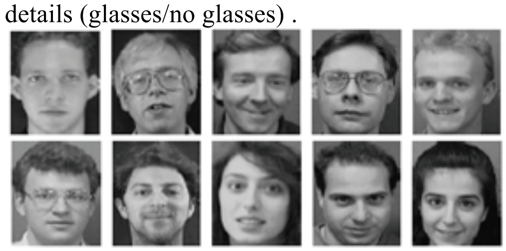

Fig.1 Face images on the ORL database.

When the difference subspace is a well-conditioned matrix, the experimental result of the proposed algorithm is the same as that of the DCV. Here, we test algorithms under the difference subspace where columns have linear correlation. We randomly selected samples 3, 4, 5, 6, 7 from each class for the training, and the remaining samples in each class for the testing. Then, we randomly selected an image with little motion using an affine transform as another image in each class. Constructing the difference subspace matrix was ill-conditioned by the above procedure because of the linear correlation among columns in this subspace.

The experiments were also repeated 10 times. The average recognition rates (arr), standard deviation (std) and LOO of the two experiments are listed in Table 2. It can be seen that the LOO of the GSR is less than that of the GSO. The recognition rate of the proposed method was superior to that of the DCV under the columns of the subspace which have linear correlation. The experimental results show that the projection vector orthogonality can be well preserved by the GSR.

Table 2: Comparison of the two methods on the ORL database (under the ill-conditioned matrix)

Method	Training sample	3	4	5	6	7
DCV	arr	0.81	0.85	0.91	0.92	0.94
	std	2.32	2.21	2.54	2.75	1.33
	LOO	13.8	49.1	29.3	59.0	28.8
The proposed method	arr	0.84	0.89	0.93	0.94	0.96
	std	2.14	1.55	2.15	2.56	1.25
	LOO	0.18	0.19	0.21	0.21	0.19

3.2 Experiments with the AR face database

The AR face database contains over 4,000 color images corresponding to 126 people's faces (70 men and 56 women)

[1]. Images feature frontal view faces with different facial expressions, illumination conditions, and occlusions (sun glasses and scarf). No restrictions on wear (clothes, glasses, etc.), make-up, hair style, and so on. We selected 1300 images of 50 males and 50 females, and each person had 13 images for testing our method. The original images are 768 x 576 pixels, and were then normalized to 128 x 128 (as shown in Fig.2).

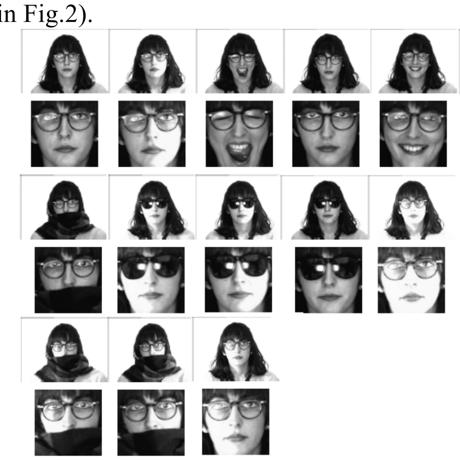

Fig.2 Original and normalized face images on the AR face database.

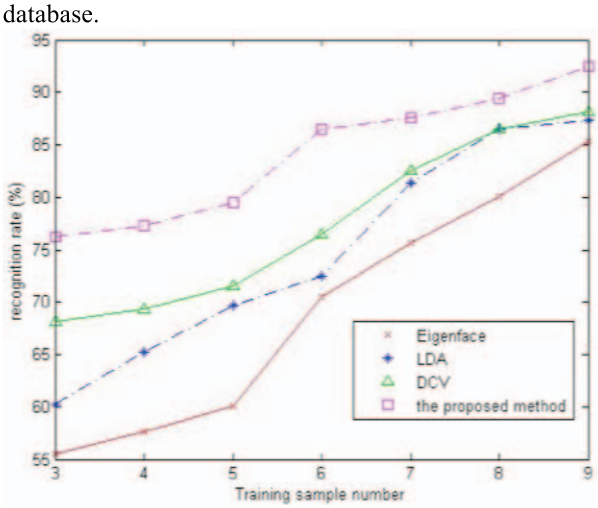

Fig.3: Recognition rate (%) comparison of different approaches on the AR database.

We designed a series of experiments to compare the performance of the proposed method, Eigenface, LDA and DCV methods under the conditions where the training sample size was varied. We randomly selected 3, 4, 5, 6, 7, 8, 9 samples from each class for the training, and the rest for the testing. In all training samples, only one image was chosen to transform by a subtle affine matrix to get a similar image. Each experiment was repeated 5 times, and the results were shown in Fig.3.

Fig.3 revealed that the proposed method was comparable to other methods in terms of recognition rate and the number of the training samples. It can be easily ascertained, with increased training number, that the proposed method obtained better recognition rate compared to other methods. The experimental results show that the DCV based on the GSR algorithm is better than the DCV when the columns of the subspace have linear correlation.

4. CONCLUSION

The classical Gram-Schmidt technique may produce a set of vectors which is far from orthogonal and sometimes the orthogonality can be lost completely. If the projective vectors obtained by the GSO are not optimal, the effectiveness of the DCV will also decrease. In this paper, we proposed an improved DCV method based on the GSR for face recognition. For obtaining an accurate projection onto the corresponding space, the orthogonal basis problem is solved usually by the Gram-Schmidt process with reorthogonalization. Our experiments showed that the performance of the proposed algorithm is the same as that of the DCV when the columns of the subspace have linear independency, while the performance of the proposed algorithm is better than that of the DCV when the columns of the subspace have linear correlation.

5. ACKNOWLEDGEMENTS

This work is supported by NSFC (No. 61170220, No. 70973046, No.60805007), Shanghai Science and Technology Development Funds (No. 10QA1407100) .

6. REFERENCES

[1] AR face database, the Computer Vision Center (CVC) , U.A.B.

[2] M.B.Gulmezoglu, V.Dzhafarov, M.Keskin, and A.Barkana, “A novel approach to isolated word recognition”, *IEEE Trans. Speech and Audio Processing*, vol. 7, no. 6, 620-628, 1999.

[3] M.B.Gulmezoglu, V.Dzhafarov, and A.Barkana, “The common vector approach and its relation to principal Component Analysis”, *IEEE Trans. Speech and Audio Processing*, vol. 9, no. 6, 655-662, 2001.

[4] H.Cevikalp, M.Wilkes and A. Barkana, “Discriminative common vectors for face recognition”, *IEEE Trans. Pattern Analysis and Machine Intelligence*, vol.27, no.1, 4-13, 2005.

[5] A.Bjorck, Solving linear least squares problems by Gram-Schmidt orthogonalization, *BIT* 7, 1-21, 1967.

[6] L.Giraud and J.Langou, “The loss of otrhogonality in the Gram-Schmidt orthogonalization process”, *Computers and Mathematics with Applications*, 50, 1069-1075, 2005.

[7] A.Bjorck, Numerical methods for least squares problems, *SIAM*, Philadelphia, PA, 1996.

[8] G.H.Golub and C.F. Vanloan, Matrix computations, Third Edition, John Hopkins University Press, *Baltimore*, MD, 1996.

[9] A.Bjorck and C.Paige, “Loss and recapture of orthogonality in the modified Gram-Schmidt algorithm”, *SIAMJ*. Matriax Anal. Appl. 13 (1), 176-190, 1992.

[10] J.R.Rice, “Experiments on Gram–Schmidt orthogonalization”, *Math. Comp*. 20, 325–328, 1966.

[11] A.Dax, “A modified Gram-Schmidt algorithm with iterative orthogonalization and column pivoting”, *Linear algebra and its applications* 310, 25-42, 2005.

[12] L.Giraud, J.Langou, and M.Rozloznik. “On the round-off error analysis of the Gram-Schmidt algorithm with reorthogonalization”. Technical Report TR/PA/02/33, CERFACS, Toulouse, France, 2002.