

An Incremental Learning Face Recognition System for Single Sample Per Person

Tao Zhu, Furao Shen and Jinxi Zhao

Abstract—Making recognition more reliable under the condition of single sample per person is a great challenge in computer vision. In this paper, we propose a subspace based face recognition system which focuses on dealing with this problem. Inspired by the Single Image Subspace (SIS) method and the concept of typical machine learning algorithms, we design an online incremental learning system which can keep learning information from input images to improve the system performance. By combining the strengths of principal angles based similarity measure, a threshold policy and a novel sample subspace updating algorithm, the task of robust face recognition is accomplished. Experimental results on AR and EYALE database are presented to demonstrate the effectiveness of the proposed method.

I. INTRODUCTION

During the past 20 years, face recognition technology has received great attentions from both the academic and industrial communities for their continuous growing needs. A great deal of approaches has been proposed and many of them are proved to be effective. While encouraging results have been obtained, most of these existed methods still suffer from the Small Sample Size (SSS) problem [1] which occurs when the number of training samples is far smaller than the dimensionality of the samples. When the training set is extremely small and only one sample per person is available, their performance is severely influenced by face appearance variations and may even fail to work.

PCA or LDA has been widely used in face recognition tasks. As a result of that, designing improved LDA-based or PCA-based algorithms may be a natural choice to overcome the SSS problem [2], [3]. However, few of them can achieve a very high accuracy under the condition of one sample per person. Geometric feature-based methods [4] or topology graph matching technologies [5] are also preferred by some researchers to solve the SSS problem. However, their drawback is obvious: once the key positions are occluded, they will fail to work. Some methods try to attack the SSS problem by employing generic learning technologies [6], [7]. However, collecting enough samples to build a generic training set is not an easy task.

In recent years, researchers begin to employ incremental learning topologies to improve the performance of face

recognition system. In nowadays, most of the existed incremental learning systems are designed to update the eigenspace of face data as new images arrive [8]. To our knowledge, few of them can automatically decide when to learn new information from an input image. In other words, they need an external observer to tell them how to prevent learning distorted information from a misclassified or non-ideal image. Moreover, few of these methods can be applied in the scenario of single sample per person.

In this paper, we mainly focus on the issue of robust incremental face recognition under the condition of one training sample per person. Inspired by the Single Image subspace (SIS) approach [9], we propose an incremental learning face recognition system. The goals of the proposed system are: (1) self-adaptively updating and adjusting training samples during learning process; (2) keeping learning new knowledge in an online environment and automatically refusing non-ideal information which would destroy the original sample structure; (3) achieving a high accuracy.

II. SIS APPROACH

In SIS [9], each $M \times N$ normalized image is divided into several small regions (sub-blocks) size of $R \times C$. To balance the robustness and efficiency, the values R and C are set to be similar and they are about 1/3 of M and N respectively. The adjacent regions share more than half of their pixels so that some organs such as eyes are more likely to reside in certain partitioned regions.

Each given sub-block A generates $m \times n$ number of synthesized samples $A_{11}, A_{12}, \dots, A_{mn}$ by shifting operation:

$$A_{ij} = A\left(i : (l + i - 1), j : (r + j - 1)\right) \quad (1)$$
$$1 \leq i \leq m, 1 \leq j \leq n$$

where m and n are parameters have to be predefined, $l = M - m + 1$ and $r = N - n + 1$.

These synthesized samples are collected to generate a subspace S_A . S_A contains all filtered images \tilde{A} generated from A through filter windows with size $m \times n$:

$$\tilde{A} = \sum_{i=1}^m \sum_{j=1}^n \omega_{ij} A_{ij} \quad (2)$$

where ω_{ij} is the coefficient of filter mask and the setting of its value is arbitrary. As a result of that, the information of each single image is greatly enriched and the features are strengthened. Every single image generates a series of subspaces (subspace group).

Tao Zhu, Furao Shen and Jinxi Zhao are with the National Key Laboratory for Novel Software Technology, and Department of Computer Science and Technology, Nanjing University, China.(email:tao144@gmail.com,{frshen,jxzhao}@nju.edu.cn)

This work was supported in part by the 973 Program 2010CB327903, the Fund of the National Natural Science Foundation of Jiangsu NSF grant BK2011567.

In recognition process, the following subspace distance metric is employed by SIS:

$$d(S_1, S_2) = \sqrt{2k - 2 \sum_{i=1}^k \cos^2 \theta_i} \quad (3)$$

where S_1 and S_2 are subspaces generated from unlabeled input image and training sample image respectively, $k = m \times n$ is the subspace dimension.

Then, the similarity scores are obtained through distance normalization. Finally, the input image is labeled by the winner class which has the maximal aggregated score.

III. THE PROPOSED SYSTEM

The shortage of SIS is that each single image is presented by a series of subspaces and once the training samples are generated they don't change any more. If we continuously generate new subspaces and insert them into the training set as new samples are obtained, the storage and computational cost of our system will greatly increase. Therefore, to achieve incremental learning, we choose to modify and update sample subspaces when new images are obtained.

Additionally, the proposed system improves SIS in these aspects: (1) a new subspace similarity measurement is adopted; (2) a novel subspace updating algorithm is proposed for incremental learning; (3) a self-adaptive learning rate is designed.

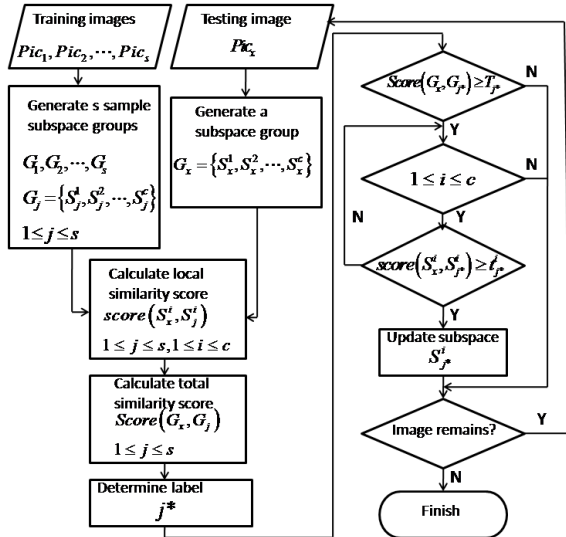


Fig. 1. Flowchart of the proposed system

As seen in Fig.1, training images of each subject are collected to generate subspaces. When a new image is input, the similarity scores are calculated. After that, we assign the label of “winner” class to the input image. The system automatically extract information from this input image to enrich the sample set by two kinds of thresholds and a subspace updating algorithm. In the following sections, the characteristics of the proposed system will be introduced and finally the whole algorithm will be summarized.

A. The Similarity Between Subspaces

When the subspaces are generated, the face recognition problem is transformed into a problem of measuring the similarity between subspaces.

A plenty of subspace similarity functions have been proposed and employing principal angles based metric may be the most common way [10], [11]. The principal angles are invariant to the order of subspaces’ basis vectors and can be computed by employing Singular Value Decomposition (SVD) [12]:

Let the columns of matrix P and Q be orthonormal basis for subspaces S_1 and S_2 respectively, and σ_i be the i -th singular values of $P^T Q$, then:

$$\cos \theta_i = \sigma_i, i = 1, \dots, k \quad (4)$$

where θ_i is the i -th principal angles between S_1 and S_2 .

In the proposed system, samples are presented as a series of subspaces have equal dimension and each subspace contains the features of one certain face region. The label of input image is determined by aggregating the subspace (local) similarity scores. In order to make the calculations more convenient, we hope the subspace similarity score to be in the interval $[0, 1]$.

Therefore, we propose a simpler subspace similarity measure based on principal angles:

$$score(S_1, S_2) = \frac{\sum_{i=1}^k \cos^2 \theta_i}{k} \quad (5)$$

Note that according to Equation (4), we got $\sum_{i=1}^k \cos^2 \theta_i = \|P^T Q\|_F^2$. The similarity score $score(S_1, S_2)$ can be calculated directly and needs no additional step to normalize.

In practice, the value of $\sum_{i=1}^k \cos^2 \theta_i$ is usually much smaller than k . To do the incremental learning judging in an online condition, a threshold policy is needed. However, if we employ Equation (3) or L_2 -Hausdorff distance [11] to measure the similarity, after normalization, the difference between similarity scores obtained by “winner” class and that obtained by the other classes is usually too little. Our similarity measure focuses on the spatial relationship between k -dimensional subspaces and the similarity scores distribute across in a wider interval. Thus, the threshold is much easier to determine.

Note that our subspace similarity score does not satisfy the “triangle inequality” property and only can be called as “non-metric measurement”. However, in complex applications such as pattern recognition, metric distance measurement in many cases are seen to fail and non-metric measurements work well [13].

B. Incremental Learning Judging

In the incremental learning process, if the similarity score $Score_{j^*}$ between input testing image Pic_x and the “winner” sample is very small, it is very likely that: (1) the classification of Pic_x is wrong; (2) though the classification of Pic_x is right, Pic_x locates on the margin rather than in the center of “winner” class j^* .

If the images with too much distorted information are learnt, the original structure of sample data may be destroyed. Thus we need a threshold policy to reject non-ideal incoming data. On the other hand, if the threshold policy is too strict, the system will seldom learn new information from input images. Therefore, the setting of thresholds should be a trade off.

In the proposed method, we propose a threshold policy contains two kinds of threshold: total threshold T_j and local threshold $t_j^i (1 \leq j \leq c)$.

Assume the label of “winner” class is j^* , total threshold T_{j^*} is applied to ensure that the input face image Pic_x is similar with the “winner” object and avoid misclassification. If the total similarity score is less than T_{j^*} , the proposed method stops incremental learning.

Local threshold $t_{j^*}^i$ is designed to ensure the structural similarity between subspaces generated from the same position of image Pic_x and object j^* in subspace transforming process. If the local similarity score is greater than $t_{j^*}^i$, we may believe that the updating operation would not destroy the original structure of training samples.

In training process, assume subspace group $G_j = (S_j^1, S_j^2, \dots, S_j^c)$ is generated to present the j -th person ($1 \leq i \leq c, 1 \leq j \leq s$). We calculate the similarity scores between every couple of subspace groups belong to different person: $Score(G_{j_1}, G_{j_2}) = \sum_{i=1}^c score(S_{j_1}^i, S_{j_2}^i)$.

The total similarity threshold of the j_1 -th object is set to be:

$$T_{j_1}' = \max(Score(G_{j_1}, G_{j_2})), j_2 = 1, 2, \dots, s, j_2 \neq j_1 \quad (6)$$

$$T_{j_1} = \max(T_{j_1}' + 1, \frac{c}{3}) \quad (7)$$

The setting of total similarity threshold mainly considers on two aspects:

- 1) $\max(Score(G_{j_1}, G_{j_2}))$ is the maximum total score between the j_1 -th object and every other object. Then, +1 operation is employed to prevent the situation that this value is too small.
- 2) $\frac{c}{3}$ is a given empirical parameter, c is also the maximum value the total similarity score can get.

We define local threshold of the i -th sample subspace belongs to the j -th object as:

$$t_{j_1}^i = \text{mean}(score(S_{j_1}^i, S_{j_2}^i)), j_2 = 1, 2, \dots, s, j_2 \neq j_1 \quad (8)$$

C. Subspaces Updating

To fulfill the incremental task, many methods prefer to directly add new samples into the existing sample set. However, it might result in a waste of storage, and the computational cost may increase greatly. We choose another direction: modifying and updating sample subspaces when new samples are obtained.

Assume S_1 is a k -dimensional sample subspace in \mathbb{R}^n :

$$S_1 = \text{span}(P), P = [p_1, p_2, \dots, p_k] \in \mathbb{R}^{n \times k}$$

p_i is S_1 's orthonormal basis vector.

Let $S_2 \in \mathbb{R}^n$ be a incoming k -dimensional subspace:

$$S_2 = \text{span}(Q), Q = [q_1, q_2, \dots, q_k] \in \mathbb{R}^{n \times k}$$

q_i is S_2 's orthonormal basis vector.

We want to adjust S_1 and generate a new sample subspace S_3 , S_3 can present the features of both S_1 and S_2 . A subspace consists of all the linear combinations of its basis vectors and can be adjusted by updating its basis.

In machine learning, the updating of vector is usually in the following form:

$$w_i^{new} = w_i^{old} + \eta_i(v_{input} - w_i^{old}) \quad (9)$$

$\eta_i \in [0, 1]$ is the learning rate, w_i is the sample vector and v_{input} is the newly input vector.

Before updating the basis vectors, we have to ensure the one-to-one correspondence between these vectors. As a result of that, we should rearrange these basis vectors of S_1 and S_2 :

$$S_1 = \text{span}(P'), P' = [p'_1, p'_2, \dots, p'_k] \in \mathbb{R}^{n \times k}$$

$$S_2 = \text{span}(Q'), Q' = [q'_1, q'_2, \dots, q'_k] \in \mathbb{R}^{n \times k}$$

where p'_i and q'_i are a pair of basis vectors, and the angle between them is the i -th principal angle θ_i between subspaces S_1 and S_2 .

According to the properties of principal angles, p'_i and q'_j satisfy:

$$(p'_i)^T q'_j = \begin{cases} \sigma_i & i = j \\ 0 & i \neq j \end{cases} \quad (10)$$

where σ_i is the i -th singular value of $P^T Q$.

Moreover, p'_i and q'_i can be determined directly through SVD ($USV = P^T Q$):

$$P' = PU \quad Q' = QV \quad (11)$$

Note that $P^T Q$ is a matrix with low dimension ($k \times k$). Thus, the cost of taking SVD on $P^T Q$ is not great.

When the one-to-one correspondence between p'_i and q'_i is satisfied, we can calculate a new basis vector w'_i and normalize it:

$$w'_i = (1 - \eta_i)p'_i + \eta_i q'_i \quad (12)$$

$$w_i = \frac{w'_i}{\|w'_i\|_2} \quad (13)$$

learning rate η_i is set in a range $[0, 1]$ and the details of this will be discussed in Section III-D.

According to Equation 10 and Equation(12), we have

$$\begin{aligned} (w'_i)^T w'_j &= ((1 - \eta_i)p'_i + \eta_i q'_i)^T ((1 - \eta_j)p'_j + \eta_j q'_j) \\ &= \begin{cases} (1 - \eta_i)^2 + 2(1 - \eta_i)\eta_i\sigma_i + \eta_i^2 & i = j \\ 0 & i \neq j \end{cases} \end{aligned} \quad (14)$$

It is no doubt that $(1 - \eta_i)^2 + 2(1 - \eta_i)\eta_i\sigma_i + \eta_i^2 > 0$.

Finally,

$$w_i^T w_j = \frac{(w'_i)^T w'_j}{\|w'_i\|_2 \|w'_j\|_2} = \begin{cases} 1 & i = j \\ 0 & i \neq j \end{cases} \quad (15)$$

In brief, given a sample subspace $S_1 = \text{span}(P)$ and an incoming subspace $S_2 = \text{span}(P)$, the updating of sample subspace is taken in following steps: (1) employ SVD on $P^T Q$; (2) rearrange the basis of S_1 and S_2 by Equation (11); (3) simultaneously adjust the k pairs of basis vectors by Equation (12) and (13).

$S_3 = \text{span}(w_1, w_2, \dots, w_k)$ is the newly updated sample subspace and w_1, w_2, \dots, w_k forms its orthonormal basis.

D. Learning Rate

Inspired by neural-gas network [15], the change of vectors can be wrote as

$$\Delta w_i = \epsilon \cdot f_i \cdot (q'_i - p'_i) \quad (16)$$

where $\epsilon \in [0, 1]$ is the step size which determines the overall extent of modification; $f_i \in [0, 1]$ accounts for the relationship of the vectors p'_i and q'_i .

In our incremental learning system, the step size is determined as $\epsilon = 1/(\text{time} + 1)$. Parameter *time* is the number of times a subspace has been updated. As *time* increases, ϵ trends to be 0.

According to the definition of principal angles, even among two very different subspaces the first and the second principal angles of them are usually very small. Considering the one-to-one correspondence between pair vectors p'_i and q'_i and the simultaneously performed adjustment between these pairs, in order to actually improve the structure of sample subspaces, the basis vector pairs in lower-ranking should be paid more attention. As a result of that, we set $f_i = e^{-\sigma_i}$.

We get the learning rate:

$$\eta_i = \frac{1}{\text{time} + 1} \cdot e^{-\sigma_i} \quad (17)$$

E. Algorithm

Based on the analysis above, we present the main process of the proposed method in Algorithm 1.

In the proposed method, all images of the same subject are presented as one single subspace group. When an unlabeled image Pic_x is input, a subspace group G_x is generated from it. After the subspace similarity measuring, Pic_x 's label is determined. When threshold standards are met, information is extracted from G_x and the sample subspaces are adjusted. Moreover, the proposed method can successfully work in an online condition.

Algorithm 1 Incremental Learning Algorithm

- 1: Preprocess on all face images.
- 2: Generate sample subspace groups G_1, G_2, \dots, G_s to present s different subjects. For each subspace group G_j ($1 \leq j \leq s$), we have $G_j = \{S_j^1, S_j^2, \dots, S_j^c\}$
- 3: Input an unlabeled image Pic_x , generate a subspaces group $G_x = \{S_x^1, S_x^2, \dots, S_x^c\}$ from it.
- 4: Calculate the local similarity score $score(S_x^i, S_j^i)$ ($1 \leq i \leq c, 1 \leq j \leq s$) by Equation(5).
- 5: The label of the testing image j^* is obtained through calculating the maximal total similarity score:

$$Score(G_x, G_j) = \sum_{i=1}^c score(S_x^i, S_j^i) \\ j^* = \arg \max_j Score(G_x, G_j) \quad (18)$$

where $1 \leq j \leq s$, s is the number of different subjects.

- 6: Incremental learning:
 - i) If $Score(G_x, G_{j^*}) < T_{j^*}$, goto Step(7).
 - ii) for $i = 1, 2, \dots, c$
 - a) If $score(S_x^i, S_{j^*}^i) < t_{j^*}^i$, goto Step (7).
 - b) Employ the i -th subspace of G_x to update and adjust the corresponding subspace of subspace group G_{j^*} .
 - 7: Goto step 3 to continue the recognition process until there is no remaining unlabeled face image.
-

IV. EXPERIMENT

A. Database and Experimental Settings

To evaluate the proposed face recognition system, two well-known face databases are applied in our experiments: AR [14] and Extended Yale (EYALE) [16], [17]. Both of these two databases have a standard cropped version which is widely used in algorithm evaluation.

For AR, we employ a subset preprocessed and provided by Martinez [14]. This subset contains 2600 color images corresponding to 100 people (50 men and 50 women). Each subject has 26 different images, we normalize these 165×120 RGB images to 66×48 gray-level images. EYALE Face Database consists of 38 subjects in 64 different illumination conditions. For our experiments, we employ the manually cropped images provided by [17]. Except the corrupted images during acquisition, there are 2414 frontal face images. We also resize these original 192×168 images to 66×48 .

We define following parameters: the normalized face image size $M \times N$, the number of subblocks c , subblock size $R \times C$ and the overlapping area size. The details of parameter settings are listed in Table I. These settings are applied in all our experiments.

The setting of $k = m \times n$ which is the dimension of generated subspaces may greatly influence the recognition results. We will evaluate its setting by the later experiments.

B. Experiment 1: Influence of Subspace Dimensions

To study the influence of the subspace dimension $k = m \times n$, we conduct experiments on the AR database. This

	AR	EYALE
$M \times N$	66×48	60×60
c	30	25
$R \times C$	18×16	20×20
overlapping area	12×8	10×10

TABLE I
PARAMETERS SETTINGS

experiment employs the first image of each person in AR for training and the remaining 25 images for testing.

The degree of information expansion is worth studying. If the subspace dimension is too small, the subspaces are not able to present enough face features of one subject. On the other hand, if the dimension is too large, the subspaces may contain too much noisy information. Learning noise is not only a waste of time, but also may destroy the construction of original training samples and reduce the recognition accuracy. As a result of that, the dimension of sample subspaces $k = m \times n$ must be determined cautiously.

Consider the amount of information delivered, Liu declares that $3 \leq m, n \leq 6$ are good choices for their approach by a series of experiments [9].

In order to determine our optimal parameter settings and verify Liu's experiment result, we design the following experiments:

Firstly, the performance of the proposed method with no incremental learning with different $m \times n$ parameter settings is evaluated. To be convenient, we let $m = n$. As shown in Table II, we can observe that: for the proposed method, the settings of $3 \leq m, n \leq 6$ may also be good choices for the proposed method with no incremental learning.

$m \times n$	2×2	3×3	4×4	5×5	6×6	7×7
Result	84.2	92.4	93.2	93.1	92.2	89.1

TABLE II
INFLUENCE OF $m \times n$ TO THE PROPOSED METHOD WITH NO INCREMENTAL LEARNING(%)

Then, we test the performance of the proposed method with incremental learning. During the incremental process, the input sequence of testing data can dramatically affect the finally performance. In order to evaluate the method rigorously, we consider two different conditions: one is that the input sequence of testing images is the same as the original order of these images in AR database; the other is that the input order of testing images is totally random.

Obviously, the first condition is more ideal. Because the proposed method can extract more useful information from the easy images with little variations before the harder images are input. However, in the real-world application, we can't control the input order of the unlabeled face images. Therefore, it is important to evaluate the performance improvement of the proposed method under the condition of random input order.

Fig.2 reports the recognition accuracies by varying values of $m \times n$. When the testing images are input in random order, we get an average result of 20 runs.

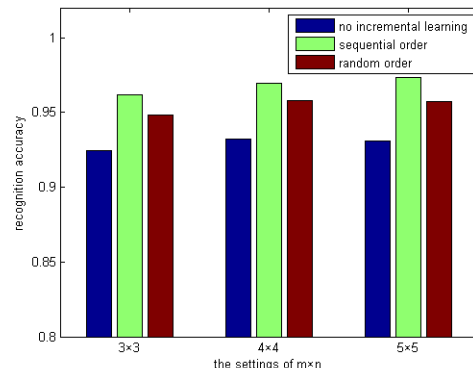


Fig. 2. Influence of $m \times n$ to the proposed method on AR grayscale (%)

Consider the accuracy and the computational cost, we let $m = n = 4$ in the following experiments.

From Fig.2, we also can find that: no matter what value the parameter $m \times n$ is, the proposed method can achieve a higher accuracy with incremental learning.

Fig.3 shows the error rate for the proposed method with and without incremental learning when $k = 4 \times 4$. The results are both got by 20 runs of random order input. It is obvious that when the proposed method incrementally learns information, the error rate declines steadily and continuously.

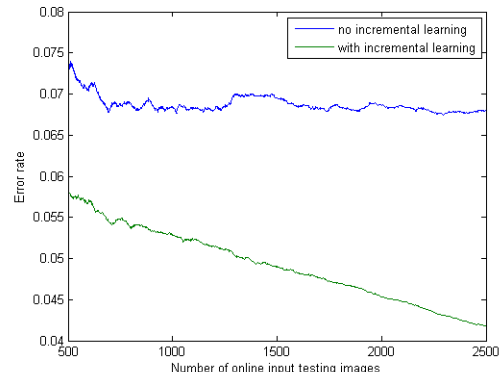


Fig. 3. Error rate as the number of input images is increased.

C. Experiment 2: Comparison with Typical Methods

In this section, we compare the proposed method with several typical methods on well-known face databases AR and Extended YALE. These experiments are all taken in the scenario of one training sample per person.

1) *Experiments on AR Database:* The first images of each subject from the AR database are employed for training and the remaining images are employed for testing.

In this experiment, we compare the proposed method with PCA, SIS [9] and Enhanced Local Binary Decisions

on Similarity (ELBDS) [18] on AR grayscale. SIS and the proposed method employ the same parameter settings listed in Table I. The PCA approach is performed as a baseline and its result reported here is the optimal one. (The optimal result of SIS reported in [9] is 94.2%.)

Method	Accuracy
PCA [9]	41.7
SIS [9]	93.3
ELBDS [18]	95.0
Ours (no incremental learning)	93.2
Ours (sequential order)	96.9
Ours (random order)	95.8

TABLE III
RECOGNITION ACCURACIES (%) ON AR GRAYSCALE IN THE SCENARIO OF SINGLE TRAINING SAMPLES PER PERSON

From the classification performance reported in Table III, we can observe that: without incremental learning, the result of the proposed method is still good. When the function of incremental learning has been introduced, the performance of the proposed method is significantly improved. Comparing with other typical methods, it has higher recognition accuracy.

2) *Experiments on EYALE Database:* In this section, we employ the first face image (which is taken under the standard lighting condition) of each subject as training set while all the other images are used for testing.

Except PCA, SIS [9] and ELBDS [18], we also compare the proposed method with Tan's method [19] which is designed for face recognition under difficult lighting conditions.

Method	Accuracy
PCA [9]	32.2
SIS [9]	99.3
ELBDS [18]	97.9
Tan's [19]	99.0
Ours (no incremental learning)	99.3
Ours (sequential order)	99.7
Ours (random order)	99.5

TABLE IV
RECOGNITION ACCURACIES (%) ON EYALE IN THE SCENARIO OF SINGLE TRAINING SAMPLE PER PERSON

The results listed in Table IV show that under the condition of illumination variation, even with no incremental learning the proposed method achieves a high accuracy (same as SIS). By learning from incoming images, the performance of the proposed method can still be improved.

V. CONCLUSION

In this paper we have presented an incremental learning face recognition system for single sample per person problem under online environments. A subspace similarity measurement suits incremental learning is proposed. By employing a threshold policy and a novel subspace updating algorithm, the proposed method can automatically extract face features

from input testing images to actually improve the classifier without recalculation. In the scenario of one training sample per person, the proposed method keeps high accuracy and is proved to outperform several typical approaches.

REFERENCES

- [1] X. Y. Tan, S. C. Chen, Z. H. Zhou and F. Zhang, "Face recognition from a single image per person: A survey," *Pattern Recognition*. vol. 39, no. 9, pp. 1725-1745, 2006.
- [2] Q. Gao, L. Zhang and D. Zhang, "Face recognition using FLDA with single training image per person," *Applied Mathematics and Computation*. vol. 205, pp. 726-735, 2008.
- [3] Y. L. Zhu, J. Liu and S. C. Chen, "Semi-random subspace method for face recognition," *Image and Vision Computing*. vol. 27, no. 9, pp. 1358-1370, 2009.
- [4] R. Brunelli and T. Poggio, "Face recognition: features versus templates," *IEEE Trans. Pattern Anal. Mach. Intell.* vol. 15, no. 10, pp. 1042-1062, 1993.
- [5] M. Lades, J. Vorbruggen, J. Buhmann, J. Lange, Malsburg von der and R. Wurtz, "Distortion invariant object recognition in the dynamic link architecture," *IEEE Trans. Comput.* vol. 42, no. 3, pp. 300-311, 1993.
- [6] Y. Su, S. G. Shan, X. L. Chen and W. Gao, "Adaptive generic learning for face recognition from a single sample per person," *Computer Vision and Pattern Recognition (CVPR)* 2010.
- [7] J. Wang, K. N. Plataniotis and J. W. Lu and A.N. Venetsanopoulos, "On solving the face recognition problem with one training sample per subject," *Pattern Recognition* vol. 39, pp. 1746-1762, 2006.
- [8] M. Cheng, Y. Y. Tang, T. P. Zhang and J. Wen, "Incremental embedding and learning in the local discriminant subspace with application to face recognition," *IEEE transactions on systems man and cybernetics part c-applications and reviews* vol. 40, no. 5, pp. 580-581, 2010.
- [9] J. Liu, S.C. Chen, Z.H. Zhou and X. Tan, "Single image subspace for face recognition," *Proceedings of the 3rd International Workshop on Analysis and Modeling of Faces and Gestures*. vol. 4778, pp. 205-219, 2007.
- [10] Q. Yamaguchi and K. Fukui and K. Maeda, "Face recognition using temporal image sequence," *Proceedings of the 3rd International Conference on Face and Gesture Recognition*. pp. 318-323, 1998.
- [11] L. W. Wang, X. W. and J. F. Feng, "Intrapersonal subspace analysis with application to adaptive bayesian face recognition," *Pattern Recognition*. vol. 4, no. 38, pp. 617-621, 2005.
- [12] A. Bjorck and G.H. Golub, "Numerical methods for computing the angles between linear subspaces," *Math. Comp.* vol. 27, pp. 579-594, 1973.
- [13] X. Y. Tan, "Towards robust face recognition from a single training sample per person use self-organizing maps," *PhD thesis, Nanjing University*. 2005.
- [14] A. M. Martinez and R. Benavente, "The AR face database," *Technical Report, CVC*. pp. 318-323, 1998.
- [15] T. M. Martinez, S. G. Berkovich and K. J. Schulten, "Neural-gas network for vector quantization and its application to time-series prediction," *IEEE Transactions on Neural Networks*. vol. 4, no. 4, pp. 558-569, 1993.
- [16] A. S. Georghiadis and P. N. Belhumeur and D. J. Kriegman, "From few to many: Illumination cone models for face recognition under variable lighting and pose," *IEEE Transactions on Pattern Analysis and Machine Intelligence*. vol. 23, no. 6, pp. 643-660, 2001.
- [17] K. Lee and J. Ho, and D.J. Kriegman, "Acquiring linear subspaces for face recognition under variable lighting," *IEEE Transactions on Pattern Analysis and Machine Intelligence*. vol. 27, no. 5, pp. 1-15, 2005.
- [18] A. P. James, "A Memory Based Face Recognition Method," *PhD thesis, Griffith School of Engineering*. November, 2008.
- [19] X.Y. Tan and B. Triggs, "Enhanced local texture feature sets for face recognition under difficult lighting conditions," *IEEE Transactions on Image Processing*. vol. 19, no. 6, pp. 1635-1650, 2010.