Single Image Subspace for Face Recognition

Jun Liu[1], Songcan Chen[1], Zhi-Hua Zhou[2], and Xiaoyang Tan[1]

[1] Department of Computer Science and Engineering,
Nanjing University of Aeronautics and Astronautics, China,
{j.liu, s.chen, x.tan}@nuaa.edu.cn
[2] National Key Laboratory for Novel Software Technology,
Nanjing University, China
zhouzh@nju.edu.cn

Abstract. Small sample size and severe facial variation are two challenging problems for face recognition. In this paper, we propose the SIS (Single Image Subspace) approach to address these two problems. To deal with the former one, we represent each single image as a subspace spanned by its synthesized (shifted) samples, and employ a newly designed subspace distance metric to measure the distance of subspaces. To deal with the latter one, we divide a face image into several regions, compute the contribution scores of the training samples based on the extracted subspaces in each region, and aggregate the scores of all the regions to yield the ultimate recognition result. Experiments on well-known face databases such as AR, Extended YALE and FERET show that the proposed approach outperforms some renowned methods not only in the scenario of *one training sample per person*, but also in the scenario of *multiple training samples per person* with significant facial variations.

1 Introduction

One of the most challenging problems for face recognition is the so-called Small Sample Size (SSS) problem [18, 25], i.e., the number of training samples is far smaller than the dimensionality of the samples. Meanwhile, the face recognition task becomes more difficult when the testing samples are subject to severe facial variations such as expression, illumination, occlusion, etc.

To deal with the SSS problem, we propose to represent each single (training, testing) image as a subspace spanned by its synthesized images. The employed synthesized images are the shifted images of the original single face image and thus can be efficiently obtained without additional computation and storage costs. To measure the distance between subspaces, we design a subspace distance metric that is applicable to subspaces with unequal dimensions. Moreover, to improve the robustness to the aforementioned facial variations, we divide a face image into regions, compute the contribution scores of the training samples based on the extracted subspaces in each region, and finally aggregate the scores of all the regions to yield the ultimate classification result. Since the proposed approach generates a subspace for each image (or a partitioned region of an image), it is named as SIS (Single Image Subspace).

Experiments on several well-known databases show that the proposed SIS approach achieves better classification performance than some renowned methods in the scenarios of both *one training sample per person* and *multiple training samples per person* with significant facial variations. In what follows, we will briefly review the related work in Section 2, propose the SIS approach in Section 3, report on experimental results in Section 4, and conclude this paper with some discussion in Section 5.

2 Related Work

In dealing with the SSS problem, the following two paradigms are often employed: 1) performing dimensionality reduction to lower the sample dimensionality, and 2) synthesizing virtual samples to enlarge the training set.

Among the many existing dimensionality reduction methods, PCA (Principal Component Analysis, Eigenfaces) [20] and LDA (Linear Discriminant Analysis, Fisherfaces) [1] are well-known and have become the *de-facto* baselines. Later advances on PCA and LDA include Bayesian Intra/Extrapersonal Classifier (BIC) [13], Discriminant Common Vectors (DCV) [4, 9], etc.

Our proposed SIS approach works along the second paradigm, i.e., synthesizing virtual samples, whose effectiveness has been verified in quite a few studies [3, 11, 16, 19, 23]. In [3], Beymer and Poggio synthesized virtual samples by incorporating prior knowledge, and yielded a classification accuracy of 82% with one real and 14 virtual images compared to 67% with only real samples on a database of 62 persons. Niyogi et al. [14] showed that incorporating prior knowledge is mathematically equivalent to introducing a regularizer in function learning, thus implicitly improving the generalization of the recognition system. In [23], Wu and Zhou enriched the information of a face image by combining the face image with its projection map, and then applied PCA to the enriched images for face recognition. They reported 3-5% higher accuracy than PCA through using 10-15% fewer eigenfaces. Martinez [11] proposed the Local Probabilistic Subspace (LPS) method. Specifically, Martinez synthesized virtual samples by perturbation and divided a face image into several regions where the eigenspace technique was applied to the generated virtual samples for classification. Good performance of LPS was reported on the AR [12] face database. In [16], Shan et al. proposed the Face-Specific Subspace (FSS) method. They synthesized virtual samples by geometric and gray-level transformation, built a subspace for every subject, and classified the testing sample by minimizing the distance from the face-specific subspace. The effectiveness of FSS was verified on face databases such as YALE B [6]. Torre et al. [19] generated virtual samples by using 150 linear and non-linear filters, and built an Oriented Component Analysis (OCA) classifier on each representation. By combining the results of the 150 OCA classifiers, they achieved good performance on the FRGC v1.0 dataset.

The synthesized samples are usually exploited for generating a subspace. There are roughly three styles for generating the subspace: 1) generating a subspace from the whole enlarged training set, e.g., [3, 11, 23], 2) generating a subspace from all the synthesized images of the same subject, and 3) generating a subspace from all the images passing through the same filter, e.g., [19]. In contrast to these past studies, we generate a subspace from each single (training, testing) image by exploiting its synthesized (shifted) images. To the best of our knowledge, such a style has not been reported in literature.

In addition to the SSS problem, severe facial variations such as expression, illumination and occlusion can often make the face recognition task especially hard [18, 25].

To deal with illumination variation, Georghiades et al. [6] proposed an illumination cone method. They exploited the fact that the set of images of a subject in fixed pose but under all possible illumination conditions is a convex cone in the space of images, and assigned to a testing image the identity of the closest approximated illumination cone. This method achieved perfect performance on the Yale Face Database B. Lee et al. [10] dwelled on how to arrange physical lighting so that the acquired images of each subject can be directly used as the basis vectors of a low-dimensional linear space. They proposed the Nine Points of Light (9PL) method with two versions: 9PL with simulated images (9PL$_{sim}$) and 9PL with real images (9PL$_{real}$), and verified their effectiveness on face databases such as Extended YALE. Like the illumination cone and 9PL methods, our SIS approach also employs linear subspace representation. Yet in contrast to illumination cone and 9PL which generate a subspace from all the training images of a subject, our SIS approach builds a subspace for each single (training, testing) image.

To deal with variations of expression and occlusion, Martinetz [11] proposed to divide a face image into several regions, and the LPS method yielded good performance on the AR face database. Tan et al. [17] partitioned a face image into several regions, and trained a Self-Organizing Map (SOM) on each region for feature extraction. The SOM-face method has been proven effective on databases such as AR. Moreover, in [18], it has been indicated that face recognition is less sensitive to facial variations such as expression and occlusion when a face image is divided into several regions that are analyzed separately. Inspired by these works, we also divide a face image into several regions in the proposed SIS approach to improve the robustness to the severe facial variations.

3 The SIS Approach

3.1 Generating Synthesized Samples

Given a face image matrix A of size $M \times N$, we generate the following $m \times n$ number of synthesized (shifted) images with size of $l \times r$:

$$\begin{aligned} A_{ij} = A(i:(l+i-1), j:(r+j-1)), \\ 1 \leq i \leq m, 1 \leq j \leq n, \end{aligned} \tag{1}$$

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

It is obvious that the shifted images can be obtained without additional computation and storage costs, since they are the shifted parts of the original single image. When $m = n = 1$, there is only one shifted image, i.e., the original face image; when m and n are relatively small, say $m = n = 3$, there are nine shifted images (illustrated in Fig. 1) that resemble the original face image visually; and when m and n are very large, say $m = M$ and $n = N$, there are $M \times N$ number of synthesized images that contain little visual information, since they reduce to points. Therefore, the values of m and n trade off the number of the synthesized images and the information delivered. We have observed in experiments (in Section 4.5) that $3 \leq m, n \leq 6$ are good choices for the proposed SIS approach.

Fig. 1. Illustration of a face image A and its nine synthesized (shifted) images A_{ij}'s with $m = n = 3$

In fact, the shifted images act as the basis images in the linear spatial filtering [8] of A in the way that:

$$\tilde{A} = \sum_{i=1}^{m} \sum_{j=1}^{n} w_{ij} A_{ij}, \tag{2}$$

where w_{ij}'s are the coefficients of a filter mask with size $m \times n$, and $\tilde{A}$ is the corresponding filtered image.

3.2 Building the Subspace

By using the synthesized images of the single image A, we build a subspace:

$$S_A = span(A_{11}, A_{12}, \ldots, A_{mn}), \tag{3}$$

or equivalently

$$S_a = span(a_{11}, a_{12}, \ldots, a_{mn}), \tag{4}$$

where $a_{ij} = vec(A_{ij})$ is obtained by sequentially concatenating the column vectors of matrix A_{ij}, and $a = vec(A)$. Then, a set of orthonormal basis vectors can be computed for S_a by techniques such as Gram-Schmidt orthogonalization [7] and Singular Value Decomposition [7].

From the viewpoint of linear spatial filtering [8], the subspace S_A in fact contains infinite number of linearly filtered images of A under all possible combinations of mask coefficients w_{ij}'s. Some images contained in S_A are illustrated in Fig. 2, from which we can observe that: 1) some images are able to reveal fine facial details such as eyes, e.g., those in the third and fourth columns; and 2) some images are less sensitive to illumination variation, e.g., those in second and third rows. Therefore, it is expected that the subspace S_A (or equivalently S_a) can be more helpful than the original single image A for face recognition.

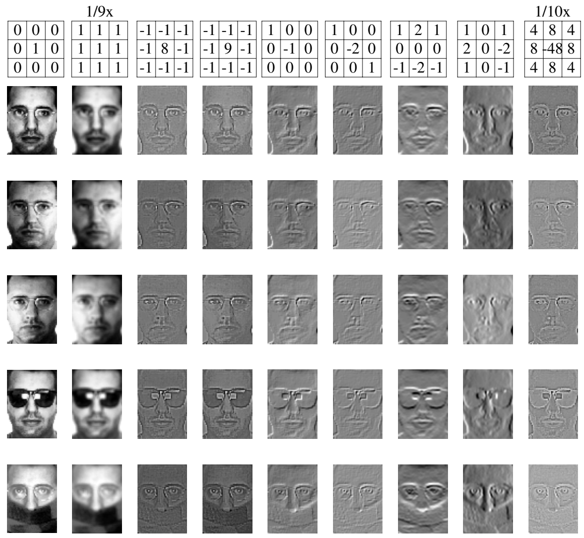


Fig. 2. Illustration of some images contained in the subspace S_A. The employed spatial filters (of size 3×3) are depicted in the first row, with the images filtered by them being shown below.

3.3 Measuring Subspace Distance

To measure the distance between subspaces, we propose the following subspace distance metric based on orthogonal projection [7]:

$$\begin{aligned} dist(S_1, S_2) =& ||PP^T - QQ^T||_F \\ =& \sqrt{k_1 + k_2 - 2||P^T Q||_F^2} \\ =& \sqrt{k_1 + k_2 - 2\sum_{i=1}^{k} cos(\theta_i)^2} \end{aligned}, \tag{5}$$

where $||.||_F$ denotes the *Frobenius* norm, T is the matrix transpose operator, S_1 and S_2 are respectively k_1- and k_2- dimensional subspaces of $\mathbb{R}^{d\times 1}$, $P \in \mathbb{R}^{d\times k_1}$ ($Q \in \mathbb{R}^{d\times k_2}$) contains a set of orthonormal basis vectors for S_1 (S_2), $k = \min(k_1, k_2)$, and θ_i's ($i = 1, 2, \ldots, k$) are the principal angles [7] between S_1 and S_2 in a non-decreasing order.

It is easy to prove that the subspace distance (5) is a metric by checking the three well-known metric properties.

Previously there are some studies on measuring the dissimilarity (or similarity) of subspaces. Golub and Van Loan [7] proposed a distance metric $dist_G(S_1, S_2) = ||PP^T - QQ^T||_2 = cos(\theta_k)$ for subspaces with equal dimensions (i.e., $k = k_1 = k_2$). Moreover, they only employed the biggest principal angle in measuring the distance. In contrast to $dist_G(S_1, S_2)$, our subspace distance metric employs the information of all the principal angles. Yamaguchi et al. [24] proposed to measure the similarity between video sequences by the smallest principal angle between subspaces. In contrast to their work which only employs the smallest principal angle, our distance metric utilizes the information of all the principal angles. Wolf and Shashua [22] proposed to measure the similarity between subspaces with equal dimensions by $\prod_{i=1}^{k}\cos(\theta_i)^2$, and they proved that this similarity makes a positive definite kernel when it is generalized to nonlinear subspaces by the kernel trick. In contrast to their similarity defined on subspaces with equal dimensions, our distance metric can deal with subspaces with unequal dimensions. Moreover, in Wolf and Shashua's method, due to the employed multiplication operator, $\prod_{i=1}^{k}\cos(\theta_i)^2$ will be dominated by some small $cos(\theta_i)$'s, i.e., even if two subspaces share $k-1$ orthonormal bases, $\prod_{i=1}^{k}\cos(\theta_i)^2$ will still be zero so long as the other basis vectors are orthogonal. It is obvious that our distance metric does not suffer from this problem. Needless to say, (5) can generate a positive definite kernel in the form of $k(S_1, S_2) = exp(-\rho dist(S_1, S_2))$ ($\rho > 0$), since Chapelle et al. [5] pointed out that $k(x, y) = exp(-\rho d(x, y))$ ($\rho > 0$) is a positive definite kernel so long as $d(x, y)$ is a metric.

3.4 Accomplishing Face Recognition

By incorporating 1) the subspace representation of a single image, 2) the subspace distance metric, and 3) the technique of dividing a face image into several regions to deal with facial variations [11, 17, 18], the SIS approach works as follows:

1) Divide a face image into c overlapping regions of size $R \times C$, with the constraints that *a*) R and C are set about 1/3 of M and N, and *b*) the adjacent regions share half of the image pixels so that some discriminative organs such as eyes are more likely to reside in certain partitioned regions;

2) Build a subspace for the i-th ($i = 1, 2 \ldots, c$) partitioned region by employing (1) and (4). Moreover, based on the subspace representation, (5) will then be used to calculate $dist_j^i$, the distance between a testing sample and the j-th training sample, and a contribution score is obtained as:

$$score_j^i = \frac{\min_{j=1}^{s} dist_j^i}{dist_j^i}, \tag{6}$$

where s is the number of training samples;

3) Assign to a testing sample the identity of the j^*-th training sample that has the maximal aggregated score as:

$$j^* = \arg\max_j \sum_{i=1}^{c} score_j^i, j = 1, 2, \ldots, s. \tag{7}$$

The time complexity of the SIS approach in classifying an unknown sample is $O(csRCm^2n^2)$.

4 Experiments

To evaluate the proposed SIS approach, we conduct extensive experiments on three well-known face databases: AR [12], Extended YALE (EYALE) [6, 10] and FERET [15].

4.1 Database Description

AR is a very challenging database that consists of over 3,200 frontal images of 126 subjects (70 men, 56 women). Each subject has 26 different images grabbed in two different sessions separated by two weeks, and in each session 13 images under severe variations in expression, illumination and occlusion were recorded. The 26 images of one subject are illustrated in Fig. 3, with the corresponding variations described in Table 1. In this paper, we use a subset provided and preprocessed by Martinez [12]. This subset contains 2,600 face images corresponding to 100 subjects (50 men, 50 women) where each subject has 26 different images under the aforesaid variations. The original resolution of these image faces is 165×120 and we resize them to 66×48.

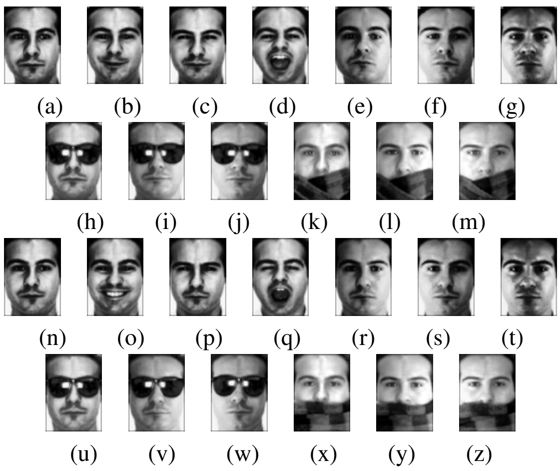


Fig. 3. Illustration of 26 images of one subject from AR.

Table 1. Variations of one subject's 26 images on AR

Session 1		Session 2	
Index	Variation(s)	Index	Variation(s)
(a)	neutral	(n)	time
(b-d)	expression	(o-q)	expression & time
(e-g)	light	(r-t)	light & time
(h)	glasses	(u)	glasses & time
(i-j)	glasses & light	(v-w)	glasses & light & time
(k)	scarves	(x)	scarves & time
(l-m)	scarves & light	(y-z)	scarves & light & time

The Extended YALE (EYALE) face database contains 2,432 frontal face images of 38 subjects under 64 different illumination conditions. The images of each subject are partitioned to five subsets according to the illumination conditions illustrated in Fig. 4. Lee et al. [10] manually cropped the images to 192×168, and we resize them to 60×50.

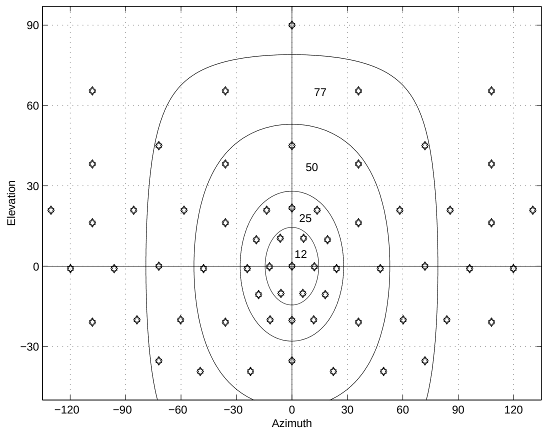


Fig. 4. The azimuth and elevation of the 64 strobes. Each annulus contains the positions of the strobes corresponding to the images of each illumination subset-Subset 1 (12^o), Subset 2 (25^o), Subset 3 (50^o), Subset 4 (77^o) Subset 5 (outer annulus) [6].

The FERET database consists of a total of 14,051 gray-scale images representing 1,199 subjects, with images containing variations in illumination, expression, pose and so on. In this paper, only frontal faces are considered. These facial images can be divide into five sets: *fa* (1,196 images), *fb* (1,195 images), *fc* (194 images), *dup I* (722 images) and *dup II* (234 images). With the eye locations provided by the FERET program, we crop the image size to 60×60.

4.2 Experimental Settings

The four parameters of SIS are set as follows: m and n are both set to 5, and R and C are set as in Table 2. Moreover, we will evaluate the influence of the four parameters in Section 4.5.

Table 2. Image size $M \times N$ and region size $R \times C$ on each database

	AR	EYALE	FERET
$M \times N$	66×48	60×50	60×60
$R \times C$	16×18	20×20	20×20

A degenerated variant of the proposed SIS approach is to directly represent each single (training, testing) face image as a subspace spanned by its shifted images, and to conduct the recognition by selecting the subspace which corresponds to a training sample and is with the minimal distance to the subspace corresponding to the testing sample. In this variant, a face image is not divided into several regions (or in other words, c, the number of partitioned regions, equals 1), and thus it is termed as SIS$_{nondiv}$. In order to study that whether the process of dividing the face image into several regions helps to improve the robustness of SIS, and whether SIS could work well without this process, we have also evaluated this variant in the experiments.

Moreover, we compare the proposed SIS approach with the following eleven face recognition methods: PCA [20], LDA [1], DCV [4, 9], BIC [13], Elastic Bunch Graph Matching (EBGM) [21], LPS [11], SOM-face [17], illumination cone [6], 9PL [10] and FSS [16]. The settings of these methods are as follows.

1) For PCA, it is a common practice to lower the dimensionality of the samples and then to employ a nearest neighbor classifier with *Euclidean* distance for classification. Since it is an open problem to select the the optimal number of employed eigenfaces, we exhaustively try all numbers of eigenfaces to report the their best performance.

2) For LDA and DCV, the samples are firstly projected to a subspace whose dimension is the number of classes minus 1, and then a nearest neighbor classifier using *Euclidean* distance is employed for classification.

3) For BIC, EBGM, LPS, SOM-face, illumination cone, 9PL and FSS, the results are directly cited from the literatures [2, 6, 10, 11, 16, 17].

4.3 One Training Sample per Person

In this subsection, we consider face recognition in the scenario of one training sample per person, and conduct the following experiments.

Experiment 1: We employ the first image of each subject from the AR database (i.e., Fig. 3 (a)) for training and the remaining images for testing. Moreover, to evaluate the performance under different facial variations, we categorize the 25 testing samples of each subject into 11 subsets according to the facial variations summarized in Table 1, and report the classification performance in Table 3. From this table, we can observe that: 1) SIS$_{nondiv}$ generally obtains higher classification performance than the holistic methods such as PCA (note that, the comparison with PCA is not fair for SIS$_{nondiv}$, since the recognition accuracies of PCA reported here are the optimal ones by trying all the numbers of projection vectors, and Martinetz [11] reported a recognition rate of less than 70% for (b-d) by using 20 eigenfaces); 2) since holistic methods are sensitive to severe facial variations such as expression, illumination and occlusion [18], it is reasonable that SIS$_{nondiv}$ achieves inferior classification performance to SIS; 3) SIS obviously outperforms PCA; and 4) compared to the recent methods such as LPS and SOM-face that are primarily designed for face recognition with one training sample per person under severe variations, SIS yields better classification performance especially for the testing samples with indexes (u) and (x) that are of severe variations in occlusion and time duration.

Experiment 2: We conduct experiments on FERET by using the *fa* set for training and the remaining sets for testing. Results are presented in Table 4, from which we can find that SIS obviously outperforms the state-of-the-art methods evaluated in the CSU

Table 3. Recognition accuracies (%) on AR in the scenario of one training sample per person (*the best performance in each case has been bolded*)

Index	Image(s)	SIS	SIS$_{nondiv}$	PCA	LPS [11]	SOM-face [17]
(b-d)		**99**	85	88	83	95
(h)		**99**	87	58	80	97
(k)		**98**	89	13	82	95
(o-q)		**86**	69	64	76	81
(u)		**96**	61	29	54	60
(x)		**90**	78	9	48	53
(e-g)		**100**	99	64	N/A	N/A
(i-j)		**96**	60	23	N/A	N/A
(l-m)		**97**	77	9	N/A	N/A
(n)		**100**	95	83	N/A	N/A
(r-t)		**99**	89	34	N/A	N/A
(v-w)		**82**	36	14	N/A	N/A
(y-z)		**84**	63	4	N/A	N/A

Face Identification Evaluation System [2] especially for the sets such as *fc*, *du I* and *du II* that contain significant facial variations.

Table 4. Recognition accuracies (%) on FERET (*the best performance in each case has been bolded*)

Method	*fb*	*fc*	*dup I*	*dup II*
SIS	**91**	**90**	**68**	**68**
PCA, MahCosine [2]	85	65	44	22
Bayesian, MAP [2]	82	37	52	32
EBGM_Standard [2]	88	40	44	22
PCA_Euclidean [2]	74	5	34	14
LDA_Euclidean [2]	61	19	38	14

Experiment 3: We further compare the performance under serious illumination variation on the EYALE face database. The first image of each subject (i.e., the image in the center of Fig. 4 with azimuth and elevation degrees of zero) is employed for training while the remaining images are used for testing. The classification rates are reported in Table 5, from which we can observe that SIS achieves higher recognition rates than PCA especially on Subsets 3, 4 and 5.

The above experiments verify that the proposed SIS approach is a good choice in the scenario of one training sample per person.

Table 5. Recognition accuracies (%) on EYALE in the scenario of one training sample per person (*the best performance in each case has been bolded*)

Method	Subset				
	1	2	3	4	5
SIS	**100**	**100**	**99**	**95**	**96**
PCA	95	91	21	4	3

4.4 Multiple Training Samples per Person

In this subsection, we consider face recognition in the scenario of multiple training samples per person and conduct the following experiments.

Experiment 4: We employ the seven non-occluded face images of each subject from the first session of the AR database (i.e., Fig. 3 (a-g)) for training and the remaining images for testing, and present the experimental results in Table 6. From this table, we can find that the proposed SIS approach achieves much higher classification accuracies than the other methods, especially when the testing images are with severe occlusions such as glasses and scarves (Table 6 (h-m, u-z)). Furthermore, comparing the results reported in Tables 3 and 6, we can find that SIS with only one training sample per person can even outperform methods such as LDA, DCV and PCA with seven training samples per person.

Table 6. Recognition accuracies (%) on AR in the scenario of multiple training samples per person (*the best performance in each case has been bolded*)

Index	Images	SIS	LDA	DCV	PCA
(h)		**99**	65	73	58
(i-j)		**100**	51	60	56
(k)		**100**	57	47	14
(l-m)		**99**	50	47	10
(n)		**100**	91	88	84
(o-q)		**99**	82	82	78
(r-t)		**99**	86	84	77
(u)		**97**	37	45	27
(v-w)		**97**	22	33	27
(x)		**92**	25	24	6
(y-z)		**95**	24	18	5

Experiment 5: We employ Subset 1 of the EYALE database for training and the remaining subsets for testing, and report the results in Table 7. From this table, we can find that: 1) SIS can obtain competitive performance to the state-of-the-art methods such as illumination cone (including Cones-attached [6], Cones-cast [6]) and 9PL (including $9PL_{real}$ [10] and $9PL_{sim}$ [10]) that are primarily designed for dealing with severe

illumination variation; and 2) SIS outperforms methods such as PCA, LDA, DCV and FSS especially on Subsets 4 and 5 that have severe illumination variation. Furthermore, comparing the results of Tables 5 and 7, we can find that SIS using only one training sample per person can even outperform methods such as LDA, DCV, PCA and FSS employing seven training samples per person.

Table 7. Recognition accuracies (%) on EYALE in the scenario of multiple training samples per person (*results in the rows marked by* * *are obtained by evaluating only the first 10 subjects, while results in the other rows are obtained by evaluating all the 38 subjects*)

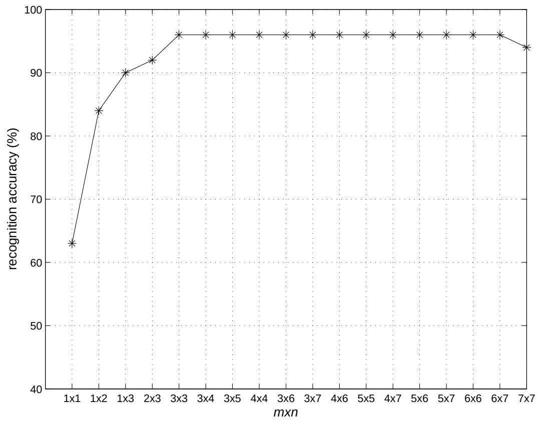
Method	Subset			
	2	3	4	5
SIS	100	100	97	99
SIS*	100	100	99	100
Cones-cast [6]*	100	100	100	N/A
9PL$_{real}$ [10]*	100	100	100	N/A
9PL$_{sim}$ [10]*	100	100	97	N/A
Cones-attached [6]*	100	100	91	N/A
FSS [16]*	100	100	87	35
LDA	100	98	37	5
DCV	100	96	32	6
PCA	90	41	6	3

The above experiments verify that the proposed SIS approach is effective in the scenario of multiple training samples per person.

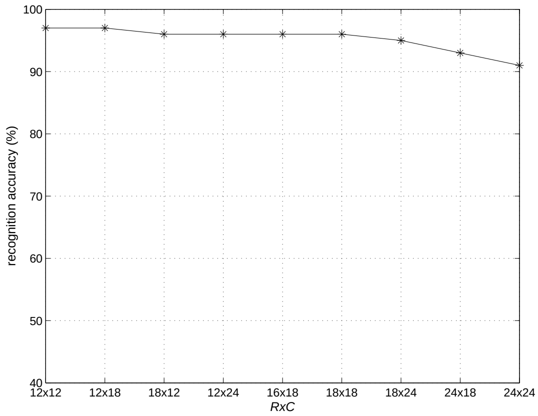
4.5 Influence of the Parameters

To study the influence of the four parameters m, n, R and C on the performance of SIS, we conduct experiments on the AR database by employing the images with index (a) for training and the images with index (u) for testing.

Experiment 6: We fix $R = 16$ and $C = 18$ and report recognition accuracies by varying values of $m \times n$ in Fig. 5 (a). First, when $m = n = 1$, for a given sub-image, we can only generate one synthesized sub-image that is just the original sub-image. In this case, SIS reduces to applying the correlation method to sub-images for classification, and only achieves a classification accuracy of 63%, which is far less than 96% by using 25 shifted sub-images. Second, when $m, n > 6$, there is a tendency towards performance drop, which may owe to the fact that little information is delivered by the synthesized sub-images with relatively big m and n, as mentioned in Section 3.1. Third, the performance curve is quite stable when the value of $m \times n$ is between 3×3 and 6×6. Note that, the recognition task here is very hard due to severe occlusion by wearing glasses, and thus it is nice to see that SIS is relatively invariant to $m \times n$ in a relatively wide range.



(a)



(b)

Fig. 5. Influence of the parameters of SIS

Experiment 7: We fix $m = 5$ and $n = 5$ and report recognition rates by varying values of $R \times C$ in Fig. 5 (b), from which we can observe that the performance curve is relatively stable when R and C are around 1/3 of M and N.

5 Conclusion and Discussion

In this paper, we propose the SIS (Single Image Subspace) approach for face recognition. First, we propose to represent each single image as a subspace spanned by its the synthesized images. Second, we design a new subspace distance metric for measuring the distance between subspaces. Third, to improve the robustness to great facial variations such as expression, illumination and occlusion, we divide a face image into several regions, compute the contribution scores of the training samples based on the extracted subspaces in each region, and aggregate the scores of all the regions to yield the ultimate recognition result.

Experiments on AR, FERET and EYALE show that SIS outperforms some renowned methods such as PCA, LDA, DCV, LPS, SOM-face, BIC, EBGM and FSS in the scenarios of both *one training sample per person* and *multiple training samples per person* with significant facial variations. Moreover, like the well-known methods such as illumination cone and 9PL, SIS yields classification accuracies close to 100% on EYALE.

In our future work we will study the theoretical justification for the proposed SIS approach. We will try to make use of kernel trick to embed the idea of SIS into traditional methods such as LDA. Moreover, we will try to exploit other synthesizing techniques in building the subspace.

Acknowledgments

The authors would like to thank the anonymous reviewers for their constructive comments that improved this paper. This work was supported in part by the National Natural Science Foundation of China under Grant Nos. 60473035 and 60635030, the National 863 project under Grant No. 2007AA01Z169, the Natural Science Foundation of Jiangsu Province under Grant Nos. BK2005122 and BK2006187, and the Innovation Foundation of NUAA under Grand No. Y0603-042.

References

1. P.N. Belhumeur, J.P. Hespanha, and D.J. Kriegman. Eigenfaces vs. Fisherfaces: Recognition using class specific linear projection. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 19(7):711–720, 1997.
2. J.R. Beveridge, D. Bolme, B.A. Draper, and M. Teixeira. The CSU face identification evaluation system: its purpose, features, and structure. *Machine Vision Application*, 16(2):128–138, 2005.
3. D. Beymer and T. Poggio. Face recognition from one example view. In *Proceedings of the 5th IEEE International Conference on Computer Vision*, pages 500–507, Washington, DC, 1995.
4. H. Cevikalp, M. Neamtu, M. Wilkes, and A. Barkana. Discriminative common vectors for face recognition. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 27(1):4–13, 2005.
5. O. Chapelle, P. Haffner, and V.N. Vapnik. Support Vector Machines for Histogram-based Image Classification. *IEEE Transactions on Neural Networks*, 10(5):1055–1064, 1999.
6. A.S. Georghiades, P.N. Belhumeur, and D.J. Kriegman. From few to many: Illumination cone models for face recognition under variable lighting and pose. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 23(6):643–660, 2001.
7. G. Golub and C. Van Loan. *Matrix Computations*. The Johns Hopkins University Press, 3rd edition, 1996.
8. R.C. Gonzalez and R.E. Woods. *Digital Image Processing*. Prentice Hall, 2nd edition, 2004.
9. J. Liu and S.C. Chen. Discriminant common vecotors versus neighbourhood components analysis and laplacianfaces: A comparative study in small sample size problem. *Image and Vision Computing*, 24(3):249–262, 2006.

10. K. Lee, J. Ho, and D.J. Kriegman. Acquiring linear subspaces for face recognition under variable lighting. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 27(5):1–15, 2005.
11. A.M. Martinez. Recognizing imprecisely localized, partially occluded and expression variant faces from a single sample per class. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 25(6):748–763, 2002.
12. A.M. Martinez and R. Benavente. The AR face database. Technical Report, CVC, 1998.
13. B. Moghaddam, C. Nastar, and A. Pentland. A bayesian similarity measure for direct image matching. In *Proceedings of the 13th International Conference on Pattern Recognition*, pages 350–358, Vienna, Austria, 1996.
14. P. Niyogi, F. Girosi, and T. Poggio. Incorporating prior information in machine learning by creating virtual examples. *Proceedings of the IEEE*, 86(11):2196–2209, 1998.
15. P.J. Phillips, H. Wechsler, J. Huang, and P. Rauss. The FERET database and evaluation procedure for face recognition algorithms. *Image and Vision Computing*, 16(5):295–306, 1998.
16. S. Shan, W. Gao, and D. Zhao. Face identification based on face-specific subspace. *International Journal of Image and System Technology*, 13(1):23–32, 2003.
17. X. Tan, S.C. Chen, Z.-H. Zhou, and F. Zhang. Recognizing partially occluded, expression variant faces from single training image per person with SOM and soft kNN ensemble. *IEEE Transactions on Neural Networks*, 16(4):875–886, 2005.
18. X. Tan, S.C. Chen, Z.-H. Zhou, and F. Zhang. Face recognition from a single image per person: A survey. *Pattern Recognition*, 39(9):1725–1745, 2006.
19. F. Torre, R. Gross, S. Baker, and V. Kumar. Representational oriented component analysis (ROCA) for face recognition with one sample image per training class. In *Proceedings of the 23rd IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, pages 266–273, San Diego, CA, 2005.
20. M. Turk and A. Pentland. Eigenfaces for recognition. *Journal of Cognitive Neuroscience*, 3(1):71–96, 1991.
21. L. Wiskott, J. Fellous, N. Kruger, and C. von der Malsburg. Face recognition by elastic bunch graph matching. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 19(7):775–779, 1997.
22. L. Wolf and A. Shashua. Learning over sets using kernel principal angles. *Journal of Machine Learning Research*, 4(6):913–931, 2004.
23. J. Wu and Z.-H. Zhou. Face recognition with one training image per person. *Pattern Recognition Letters*, 23(14):1711–1719, 2002.
24. Q. Yamaguchi, K. Fukui, and K. Maeda. Face recognition using temporal image sequence. In *Proceedings of the 3rd International Conference on Face & Gesture Recognition*, pages 318–323, Washington, DC, 1998.
25. W. Zhao, R. Chellappa, P.J. Phillips, and A. Rosenfeld. Face recognition: A literature survey. *ACM Computing Survey*, 35(4):399–458, 2003.