

Semi-Random Subspace Method for Face Recognition

Yulian Zhu^{1,2}, Jun Liu¹ and Songcan Chen^{*1,2}

1. Dept. of Computer Science & Engineering, Nanjing University of Aeronautics & Astronautics

Nanjing, 210016, P.R. China

2. State Key Lab. for Novel Software Technology, Nanjing University, P.R. China

Abstract: The small sample size (SSS) and the sensitivity to variations such as illumination, expression and occlusion are two challenging problems in face recognition. In this paper, we propose a novel method, called semi-random subspace (Semi-RS), to simultaneously address the two problems. Different from the traditional random subspace method (RSM) which samples features from the whole pattern feature set in a completely random way, the proposed Semi-RS randomly samples features on *each local region* (or a sub-image) partitioned from the original face image. More specifically, we first divide a face image into several sub-images in a deterministic way, then construct a set of base classifiers on different randomly sampled feature sets from each sub-image set, and finally combine all base classifiers for the final decision. Experimental results on five face databases (AR, Extended YALE, FERET, Yale and ORL) show that the proposed Semi-RS method is effective, relatively robust to illumination and occlusion, etc., and also suitable to slight variations in pose angle and the scenario of one training sample per person. In addition, kappa-error diagram, which is used to analyze the diversity of algorithm, reveals that Semi-RS constructs more diverse base classifiers than other methods, and also explains why Semi-RS can yield better performance than RSM and V-SpPCA.

Keywords: Random Subspace Method (RSM); Semi-Random Subspace Method (Semi-RS); Recognition Robustness; Small Sample Size (SSS); Sub-Image Method; Face Recognition; Kappa-Error Diagram.

1. Introduction:

Face recognition, as an issue of great interest and challenge in the area of computer vision, pattern recognition, etc., has been developed and widely studied over the past few decades. Among the existing methods, the subspace method [1, 2] is one of the most successful and well-studied techniques to face recognition. However, they still suffer from two problems: 1) the small sample size (SSS) problem [3], i.e. the number of training samples is far smaller than the dimensionality of the sample, which will incur the

* Corresponding author: Email: s.chen@nuaa.edu.cn

curse of the dimensionality; 2) the recognition performance is severely sensitive to variations such as illumination, expression and occlusion. In this paper, we try to solve these two problems.

Currently, the mainstream method to deal with the SSS problem is performing dimensionality reduction to lower the sample dimensionality to avoid the curse of dimensionality. Among many existing dimensionality reduction methods, Principal Component Analysis (PCA) [4] and Linear Discriminant Analysis (LDA) [23] are most well known and have become the baselines in face recognition. PCA, as an unsupervised method, is designed to perform dimensionality reduction by maximizing the variance over all the samples. LDA tries to find a set of projection basis by maximizing the between-class scatter matrix and minimizing the within-class scatter matrix in the reduced subspace. However, the SSS leads to singularity of the within-class scatter matrix in LDA. Fisherfaces [5] method mitigates the shortcoming through first performing PCA and then LDA in PCA subspace. In addition, the random subspace method (RSM) [6] is also an effective technique for overcoming the SSS problem. Different from PCA and LDA that try to find an optimal linear projection by solving certain optimization function, RSM modifies the training set by randomly sampling different smaller feature subsets from the whole original feature set, thus reduces the dimensionality of training sample and the SSS problem.

Recently, constructing random subspace in face image space has been received significant attention. Wang and Tang [7-9] first applied RSM to face recognition. They have projected face images into the PCA subspace of training image set and then performed random sampling on the PCA subspace. In [7,8], the random feature subspace is composed of two parts: fixed N_0 features associated with the N_0 largest Eigenvalues and N_1 features randomly selected from these features associated with the rest of Eigenvalues. The method achieves higher classification accuracy. Subsequently they proposed a random mixture model [9]. In [9], Wang and Tang have first randomly sampled from the whole original subspace, then clustered the intrapersonal difference, and finally computed the local intrapersonal subspaces on each random sampled feature subspace to construct base classifiers. In [11], Zhang and Jia have constructed an optimal random subspace dimensionality for discriminant analysis and further constructed a random discriminant analysis (RDA) by applying Fisher and DLDA [12], respectively, to the principal subspace of the within-class and between-class matrices on each random subspace. Moreover, Nitesh and Chawla [10] have showed the effectiveness of RSM by directly performing a complete random sampling from the PCA subspace of training set. Although these algorithms have achieved successes to a great extent, they failed to deal with well the sensitivity to larger variations such as illuminant, expression and so on. One of the

main reasons is that these methods put more emphasis on the global rather than local extraction of features that plays an important role in the robust recognition with high accuracy.

To extract more local features and to effectively deal with sensitivity to such variations as lighting, occlusion etc., researchers have successively proposed face recognition methods based on modular or sub-image [13-15, 21]. Their common point is to partition a whole face image into several smaller sub-images and then to extract local features from each sub-image respectively. Pentland et al. generalized Eigenfaces to face local feature regions and proposed Modular Eigenfaces [21]. In Modular Eigenfaces, Eigenanalysis is performed on local facial features including eyes, mouth and nose regions respectively, then the similarities between the corresponding local features of two images are calculated, finally, a sum of all these similarities forms a global similarity score of the two images. Gottumukkal has presented a modular PCA (mPCA) method [13], which enhances the robustness in illumination, expression to some extent by extracting and using local information of face images. In mPCA, firstly, a face image is divided into several smaller equally-sized sub-images, then the whole set consisting of all these sub-images is viewed as a new training set, and finally a single PCA is performed on the training set. Almost at the same time, Chen and Zhu have proposed SpPCA (sub-pattern PCA) [14]. Different from mPCA, SpPCA extracts local features separately from each sub-pattern set consisting of all the sub-patterns with the common positions or attributes, and concatenates all locally extracted features together to form a global feature vector for the original whole pattern. Extensive experiments have verified the effectiveness of SpPCA. Instead of the global concatenation on local features in [14], Aw-SpPCA in [15] uses each set of local extracted features to directly classify the unseen pattern or face. Firstly, it constructs a base classifier on each sub-image set, then classifies the corresponding (sharing the common positions with training sub-image set) sub-image of unseen image, finally fuses all classification results for the ultimate decision by using weighted majority voting rule, where the weight values denote different contributions made by different sub-images. This method achieves higher robustness on face database such as AR, Yale.

In this paper, inspired by the successes of both sub-image methods [13-15, 21] and RSM, we propose a semi-random subspace (Semi-RS) method to overcome both SSS and non-robustness for face recognition. By Semi-RS here, we mean that random sampling on *each local region* (or a sub-image) partitioned from the whole image in a deterministic way rather than random *global* sampling in RSM is performed. More specifically, we firstly divide a face image into several sub-images in a deterministic way, then construct a set of base classifiers on different feature subsets which are randomly sampled from each sub-images set,

and finally combine all base classifiers for the final decision using a combination rule. Compared with sub-pattern methods and RSM, on the one hand, from the viewpoint of sub-pattern methods, Semi-RS constructs multiple classifiers on different distribution models, so more diversity between classifiers is gained. On the other hand, from the viewpoint of RSM, Semi-RS makes use of the face's local structure information so that it can efficiently recognize the face with occlusion, illumination and expression changes. As a result, Semi-RS fuses the advantages of both the RSM and sub-image methods. We conduct experiments on five face databases (AR [16], Extended YALE [17], FERET, Yale and ORL) and results show that the proposed Semi-RS method is effective, robust to illumination, occlusion, etc., and also suitable for slight variation in pose and the scenario of one training sample per person. In addition, considering the importance of diversity in classifier ensemble, we analyze the diversities of different methods using kappa-error diagram and the kappa diagrams reveal that Semi-RS can construct more diverse base classifiers than other methods do.

The remaining parts of this paper are organized as follows. In Section 2, a brief review about the random subspace method (RSM) and sub-image method is given. In Section 3, the Semi-RS for face recognition method is formulated. In Section 4, extensive experimental comparisons carried out on five face databases are reported. In Section 5, kappa-error diagrams are plotted to illustrate the diversities of different ensemble methods. Finally, a conclusion is drawn in Section 6.

2. A Brief Review of Random Subspace Method and V-SpPCA

2.1 Random Subspace Method (RSM)

The random subspace method (RSM) [6], introduced by Ho, was proposed for constructing decision forests. RSM randomly samples a set of low dimensionality subspaces from the whole original high dimensional features space and then constructs a classifier on each smaller subspace, and finally applies a combination rule for the final decision. RSM is a very simple and popular ensemble construction method. However, sampling features directly and randomly from the whole original higher dimensional face image subspace is not practical. First, random sampling from original image subspace is time-consuming. In order to ensure both relatively sufficient sampling of RSM from high dimensional face space (or avoidance of sparse sampling) and enough good recognition accuracy, a number of feature subspaces are needed, as a result, leading to more computation. Second, random sampling from original image subspace is not helpful to deal with SSS problem because in order to gain good accuracy of component classifiers, the random

sampling rate usually needs sufficiently high, as a result, the sampled attribute or feature subset is still high dimension. Third, sampling randomly from an original image often breakdowns inherent spatial local relationship among pixels within the image but it is exactly such a locality that plays an important role in robust and high accuracy recognition. Therefore in implementation, we also adopt the same strategy as [7,8,10], i.e., first projecting all the original face images into low dimensional subspace by PCA and then performing the random sampling in the low dimensional PCA subspace.

2.2 Sub-Pattern PCA Based on Vote Rule (V-SpPCA)

SpPCA [13], as one of the most popular sub-pattern methods, can locally extract features from the whole pattern, but it is still not robust to noise mainly because SpPCA concatenates all the sub-pattern features extracted by PCA into a global features. In this paper, we improve SpPCA and propose a new SpPCA method based on majority voting rule (called V-SpPCA). V-SpPCA inherits the advantages of SpPCA as follows: it 1) extracts more local information from the original face image; 2) is helpful to deal with high dimensional image data and SSS problem by performing feature extraction in sub-pattern sets; 3) can reduce the computational complexity by performing in parallel way. But differs from the SpPCA, instead of constructing a single optimal classifier for the whole pattern, V-SpPCA constructs multiple classifiers by building a base classifier in each sub-pattern set respectively, and then combines these base classifiers for final decision. However, different from Aw-SpPCA which adopts the weighted majority voting rule, V-SpPCA uses the unweighted majority voting rule to combine the base classifiers. Compared to Aw-SpPCA, V-SpPCA is not only simple (avoiding computing the weights of all the sub-patterns) but also keeps comparable recognition accuracy. The V-SpPCA algorithm can be briefly described as follows:

Step1. Partition each training pattern into several equal-sized small sub-patterns and collect all sub-patterns sharing same original feature bases to form sub-pattern set T_i ($i = 1 \dots L$), where T_i is the i th sub-pattern training set and L is the number of sub-patterns.

Step2. Perform PCA on each sub-pattern set T_i and obtain the corresponding projection vectors Φ_i .

Step3. Classify each sub-pattern of unknown pattern Z . First partitioning Z into (Z_1, Z_2, \dots, Z_L) and then extracting sub-features Y_i from sub-pattern Z_i by $Y_i = \Phi_i^T Z_i$. Finally, classifying each Z_i ($i=1 \dots L$) with nearest neighbor classifier.

Step4. Combine all classification results using majority voting rule to for final decision.

3 Semi-Random Subspace Method

3.1 Proposed Semi-RS

Face recognition is a complex pattern recognition problem, especially when a face image set contains significant complex intrapersonal variations. Therefore, the traditional global linear subspace methods (PCA and LDA) which are based on a single Gaussian model, fail to deliver good performance. In other words, it is difficult to find a single optimal classifier that can be robust enough to obtain good classification performance. To solve this problem, face recognition methods based on RSM [6] have successively been proposed [7-11]. Their common point is to construct multiple classifiers by using a set of feature subsets randomly sampled from the original feature set and to combine these base classifiers for final decision. Since the generalization ability of the combination of multiple classifiers could be significantly better than that of a single classifier [31], these methods in [7-11] have achieved successes to a great extent. However, the RSM-based face recognition methods do not work well under the condition of varying illumination and expression, etc., since they perform random sampling in the whole image space (or PCA subspace). Under these conditions, the sampled features will vary considerably from those of the training images; hence it is not quite easy to recognize them correctly. In fact, in the real world, when there is variation in the pose, expression or illumination, only some of the face regions will vary and the rest of the regions will almost keep same as a normal image [12]. Hence the sampled features from these regions which are not affected by varying pose, expression or illumination, will closely match with the features of the same individuals face regions under normal conditions. Therefore it is expected that improved recognition accuracy with high robustness can be obtained by extracting local features from face local regions respectively. It is based on the above idea that the module or sub-images based face recognition methods have successively been proposed to boost the recognition robustness. These methods first divide an image into several smaller sub-images and then extract features from the sub-images respectively. The effectiveness of the sub-image methods has been proved through a mass of experiments. In this paper, inspired by the successes of sub-images methods and RSM, we propose a novel method, called Semi-RS, for face recognition. Semi-RS performs a random sampling on *each local region* (or a sub-image) partitioned from the whole image in a deterministic way, hence it fused advantages of the RSM and sub-image methods. Here, we take PCA as an example to describe the Semi-RS, and it is worthwhile to point out that other subspace methods such as CCA [24], NMF [25] etc. can also be incorporated into

Semi-RS, which deserves our future study. The proposed Semi-RS algorithm is described in Fig.1.

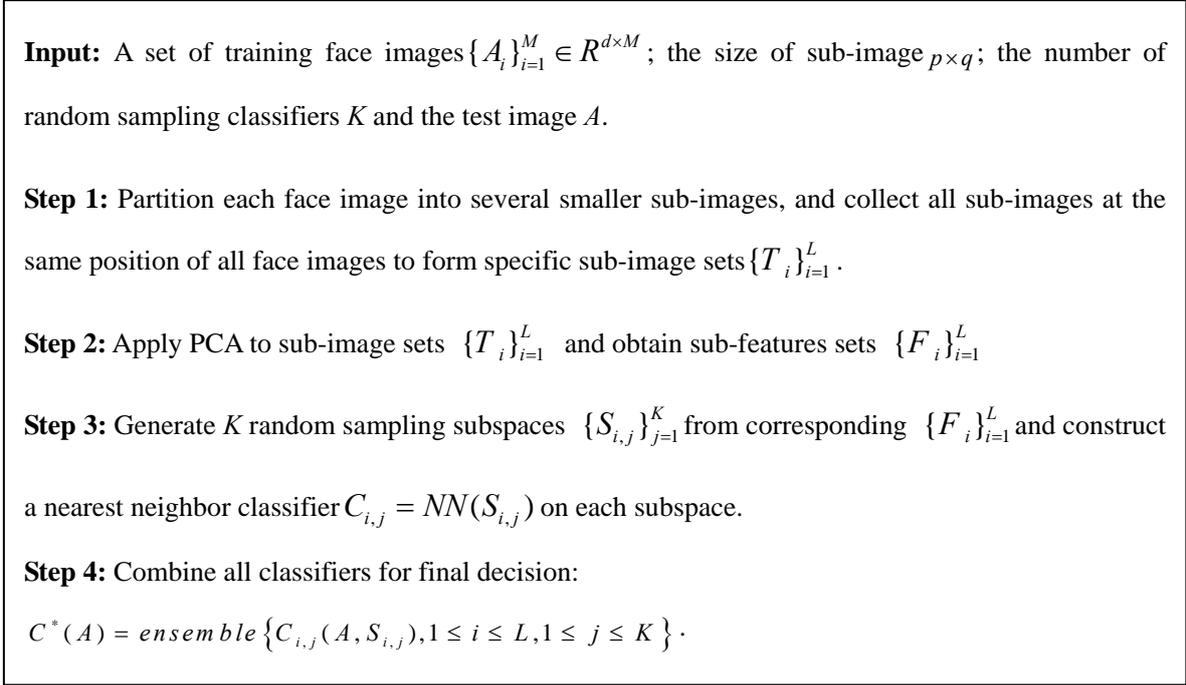


Fig.1. The proposed Semi-RS algorithm

In the first step, we divide each face image in the training set into a set of equally-sized ($p \times q$) sub-images in a non-overlapping, and hence L ($L=d/(p \times q)$) sub-images are generated for each face image. And then transform each sub-image into a column vector with dimensionality of $p \times q$ and finally collect all vectors with the common attributes of all images to form a specific sub-image's training set, in this way, L separate sub-patterns training set can be formed $\{T_i\}_{i=1}^L$. This partition process is illustrated in Fig.2.

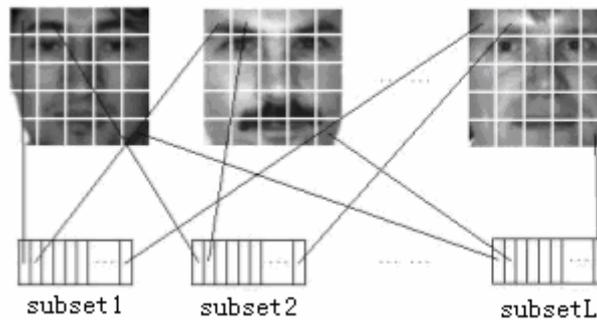


Fig.2. Construction of sub-image set

After multiple base classifiers are generated, we need to combine these base classifiers to construct an ensemble for final decision. There are many different ways to combine multiple base classifiers. In this paper, we utilize the hierarchical and parallel structures to combine them, respectively. The ensemble structure of the hierarchical Semi-RS (HSemi-RS) and parallel Semi-RS (PSemi-RS) are illustrated in Fig.

3.

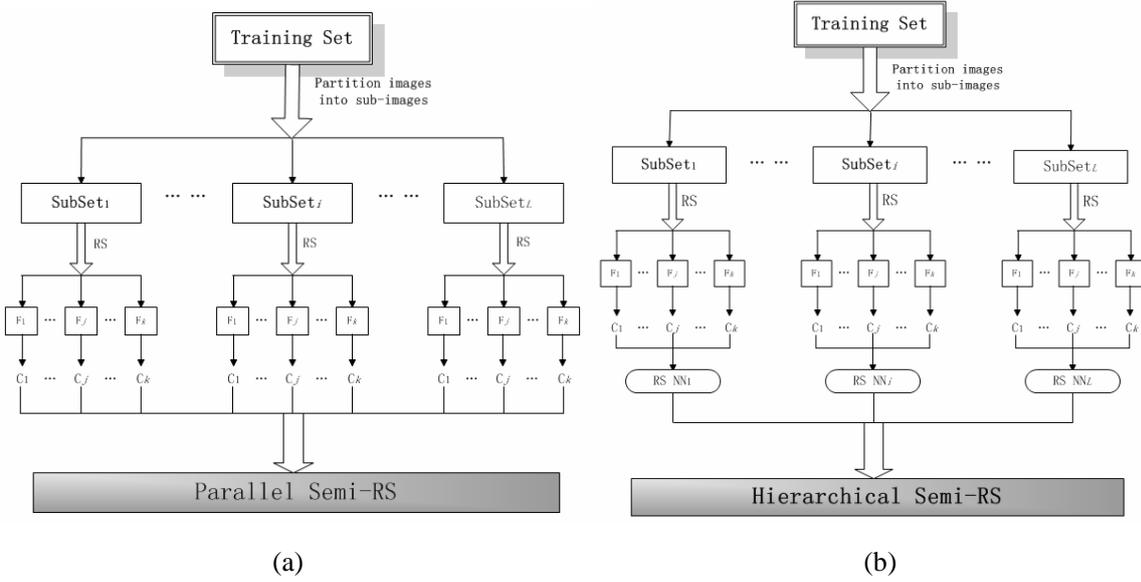


Fig.3 Structures of ensemble. (a) the parallel Semi-RS and (b) the hierarchical Semi-RS

For the PSemi-RS, when an unknown image A is given, we first divide the image into several sub-images A_i ($i=1 \dots L$) and recognize i th ($i=1 \dots L$) sub-image A_i using K weak nearest neighbor classifiers C_{ij} ($j=1 \dots K$), which are constructed on the subset of feature by randomly sampling from i th sub-feature set as described in step3. Thus, $L \times K$ weak classifiers are generated for all sub-images A_i ($i=1 \dots L$). And then, we use combination rule to construct the destination classifier. In this paper, we use majority voting rule as combination rule. If the weak base classifier $C_{ij} \in \{0,1\}$, we can represent the PSemi-RS as follows:

$$C^*(A) = \text{sgn} \left\{ \sum_{i,j} C_{ij} - \frac{KL-1}{2} \right\} \quad (1)$$

For the HSemi-RS, a double-ensemble is executed. First HSemi-RS performs a classifier combination within each sub-image to classify corresponding sub-image A_i ($i=1 \dots L$), since there are L sub-images partitioned from the image A , L classification results are obtained; and then uses the majority voting rule to combine the L classification results to classify the whole unknown images. Accordingly, when $C_{ij} \in \{0,1\}$, we can represent the double-ensemble decision as

$$C^*(A) = \text{sgn} \left\{ \sum_j \text{sgn} \left[\sum_i C_{ij} - \frac{K-1}{2} \right] - \frac{L-1}{2} \right\} \quad (2)$$

Semi-RS can generate more diverse classifiers than V-SpPCA and RS (more detailed discussion can see Section 5), so Semi-RS should outperform the two individual. In addition, in order to more sufficiently utilize the diversity, we prefer to combine all generated classifiers in parallel way as shown in Fig.3 (a).

The experiments show that the performance of PSemi-RS outperforms that of the HSemi-RS. Similar conclusion can be drawn that the parallel structure outperforms the hierarchical structure in [18].

3.2 Connection to Related Studies

1) *To refs [7- 9, 11]*: Refs [7, 8] are known as the first two published papers that apply RSM to face recognition. In [7, 8], multiple LDA classifiers are constructed on different incomplete random feature subspaces, which are composed of the first fixed N_0 dimensions and the N_1 dimensions sampled randomly from the remaining dimensions. In [9], a random mixture model was used to construct an ensemble classifier on the sampled random subspace to improve classification performance of both the Bayes and LDA subspace analysis. In [11], Zhang and Jia have constructed a multiple classifier ensemble by combining both Fishfaces and DLDA subspace to boost performance. By analyzing these methods, we find that: 1) they randomly sample from either the whole pattern subspace or PCA subspace, so the sampled features put more emphasis on the global features of images than the local structure information of face. Although [7, 8, 11] use the local features including texture, shape and Gabor wavelet, the extraction of features borrows some other algorithms rather than the methods themselves they proposed respectively; 2) they are just based on supervised learning algorithms (Bayes, LDA or NLDA) and it is basically difficult to generalize them to other unsupervised and semi-supervised subspace methods. Compared with them, Semi-RS has its advantages in the following aspects: 1) random sampling on sub-images of the whole image sufficiently considers locality of a face and thus local information between pixels can be extracted; 2) it is easy to extend to other subspace methods such as LDA[23] and NMF[25] etc. In order to demonstrate the extensibility of Semi-RS, in Section 4.7, we use LDA as a substitute for PCA to extract the local features and then give preliminary comparison with exist RSM methods [7, 8, 11].

2) *To V-SpPCA and ref [10]*: Semi-RS performs random sampling on each sub-image partitioned from the whole original image, so it is a combination of both V-SpPCA and random subspace method. In fact, Nitesh's random sampling in [10] (we call it as Nitesh'RS) and V-SpPCA can be regarded as two special cases of Semi-RS because Semi-RS will be reduced to Nitesh'RS when the number of sub-image of the whole pattern is set to 1 (i.e. no partition is performed on the whole image), while it will turn into V-SpPCA when random sampling rate in each sub-image is set to 100%. But compared with both Nitesh'RS and V-SpPCA, Semi-RS has its own advantages: on the one hand, from the viewpoint of V-SpPCA, Semi-RS constructs multiple classifiers on different distribution models, so more diversity

between classifiers is gained; on the other hand, from the viewpoint of RS, Semi-RS makes use of the face’s local structure information so that it can relatively efficiently recognize the faces with occlusion, illumination and expression changes, as shown in the following experiments.

4. Experiments

In order to evaluate the performance of the proposed Semi-RS method, we conduct experiments on five benchmark face images databases: AR [16], Extended YALE [17], FERET [19], Yale [26] and ORL [27]. Considering the specific characteristics of these five databases, the AR database is employed to test the performance of Semi-RS under variations of over time, facial expression, illumination and occlusion conditions. The Extended YALE database is utilized to evaluate the performance under severe variations of illumination. The FERET database is used to test the performance in the scenario of only one training sample per person. The Yale database is utilized to examine the stability using different training sample set. And the ORL is used to test the robustness of Semi-RS to small pose variation. Since V-SpPCA can obtain both higher accuracy recognition and higher robustness than SpPCA [14] can do, in order to show the superiority of Semi-RS in recognition performance, we only compare Semi-RS with V-SpPCA in the following experiments. In addition, in order to further evaluate the performance of Semi-RS under different distance metrics, we use two distance metrics: Euclidean and Mahcosine [20] which are the two mostly widely employed metrics in face recognition. Finally, in all the experiments, the underlying classifier is a Nearest Neighbor classifier.

Table 1. Parameters settings for Semi-RS

DataSet	Size of sub-image	Random rate
AR	6×6	0.5
Extended YALE	6×7	0.5
FERET	6×6	0.5
YALE	4×4	0.5
ORL	4×4	0.5

4.1 Experimental settings

We compare the proposed Semi-RS approach with the following face recognition methods: Eigenface, M-Eigenface [21], Fisherfaces, Nitesh’RS, V-SpPCA, Aw-SpPCA, RDA [11], R_LDA [7,8] and SVD perturbation [22]. The settings of these methods are as follows:

- 1) For Semi-RS, there are three important parameters need to be set: the size (length and width) of sub-image and random sampling rate in each sub-image. Table 1 lists their setting and a further discussion of their influence on recognition performance is given in Section 4.8
- 2) For Fisherfaces, the face images are firstly projected to PCA subspace whose dimension is the number of class minus 1, and then a nearest neighbor classifier using different distance metric is used for classification.
- 3) For SVD perturbation, we adopt the same parameters as in [22], i.e. the control parameter is set to 0.25, perturbation parameter to 1.5 and the energy to 0.95.
- 4) For Nitesh'RS in [10], we report the optimal the recognition accuracies by trying the random rate from 10% to 100% by interval 10%.
- 5) For V-SpPCA and Aw-SpPCA, the same size of sub-images as shown in Table 1 are adopted and in each sub-image set, all features extracted by PCA or LDA are used to construct the nearest neighbor classifier.
- 6) For Eigenface method, we keep the first M Eigenfaces which retain 98% energy and then a nearest neighbor classifier is utilized for classification.
- 7) For RDA in [11], the number of sampled features is set to $C-1$ (C is the number of the class) according to the experimental setting in [11].
- 8) For R_LDA in [7, 8], the numbers of fixed features and of random features are changed from 25 to 350 with increment of 25, respectively, and the optimal recognition accuracies are reported.
- 9) For M-Eigenface in [32], we determine manually each feature region (including eyes, nose and mouth) for each image and keep the first M Eigeneyes, Eigennose and Eigenmouth which retain 98% energy, respectively. The similarities between the corresponding features of two images are calculated according to their projections, and a sum of all these similarities of all corresponding features is formed as a global similarity score of the two images. The used similarity measure is as follows [32]:

$$S = \sqrt{1 - \frac{\|A - B\|}{\|A\| + \|B\|}} \quad (3)$$

For RSM-based methods (Nitesh'RS, Semi-RS, RDA and R-LDA), we repeat independently each experiment 10 times and report the averaged accuracies.

4.2 Experiments on the AR Database

The AR face database is a very challenging database that contains over 3,200 frontal face images of 126 different individual (76 men and 50 women). Each individual has 26 different images taken in two

different sessions separated by two weeks intervals and each session consists of 13 faces with different facial expressions, illumination conditions and occlusions. In our experiments, we use the subset of AR database provided and preprocessed by Martinez [16]. This subset contains 2,600 face images corresponding to 100 subjects (50 men, 50 women) and provides 26 shots for each person. The original resolution of these face images is 165×120 and here resized to 66×48 . Fig.4 shows all samples of one person in AR. In both sessions, the detail information of the images are from left to right: 1st neutral expression, 2nd to 4th expression variation, 5th to 7th illumination, 8th to 10th sunglass occlusion and 11th to 13th scarf occlusion. In the experiments, the first seven images in session1 are selected for training and the rest images which are divided into seven subsets according to the variation category (Please see Table 2) are utilized to examine the performance of proposed methods under variance conditions.

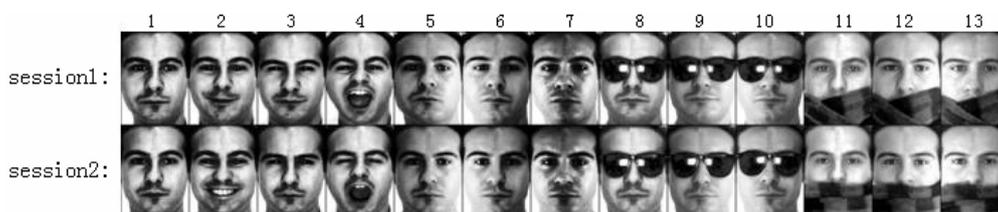


Fig.4. Examples images in AR face database

Table 2. The test Sets in AR database

Test Set	Train Set	Test samples
AR77	1 st -7 th in Session1	1 st -7 th in Session2
AR73Exp		2 nd -4 th in Session2
AR73Illu		5 th -7 th in Session2
AR73SungS1		8 th -10 th in Session1
AR73SungS2		8 th -10 th in Session2
AR73ScarfS1		11 th -13 th in Session1
AR73ScarfS2		11 th -13 th in Session2

4.2.1 Variation of expression and illumination

In this experiment, AR73Exp and AR73Illu are used for testing the performance of Semi-RS under the expression and illumination variations. Table 3 lists the performance comparison of Semi-RS with other methods using two distance metrics. From the Table, we can get the following observations: 1) compared to Eigenfaces, Fisherfaces and Nitesh'RS, the other five methods based on local feature gain more satisfactory classification results, which can attribute to the fact that the images in AR are more various in expressions and illuminations, thus it is more reasonable to use local feature to limit the effect of

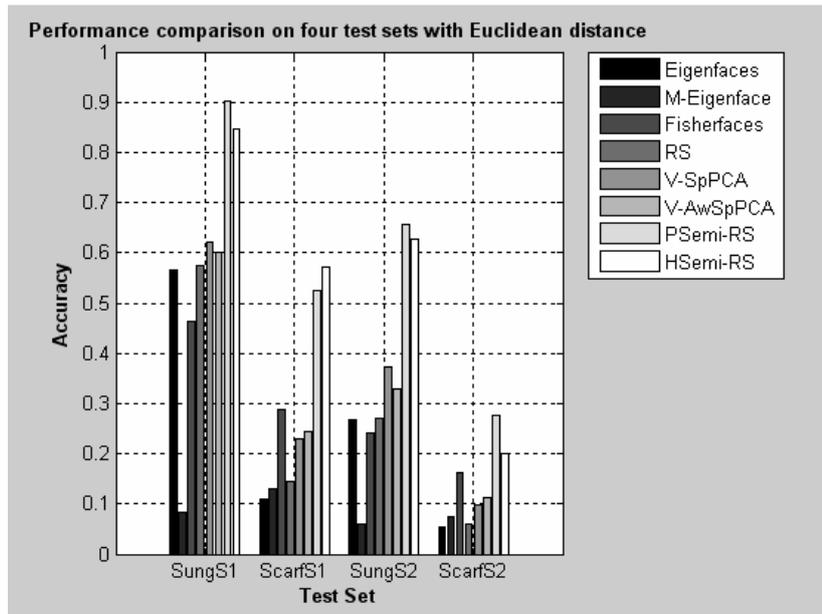
lights or expressions; 2) among methods based on local feature, PSemi-RS gains the best classification accuracy; especially under variations of illumination (AR73Illu test set), PSemi-RS and HSemi-RS recognize correctly all test images; 3) the two Semi-RSs (PSemi-RS and HSemi-RS) exhibit their effectiveness and superiority to other compared methods with respect to the two different (Euclidean and Mahcosine) distance metrics; 4) PSemi-RS outperforms HSemi-RS, the main reason is that PSemi-RS utilize diversity more sufficiently than HSemi-RS does.

Table 3. Recognition accuracies (%) on AR73Exp and AR73Illu (the best performance in each case is bolded)

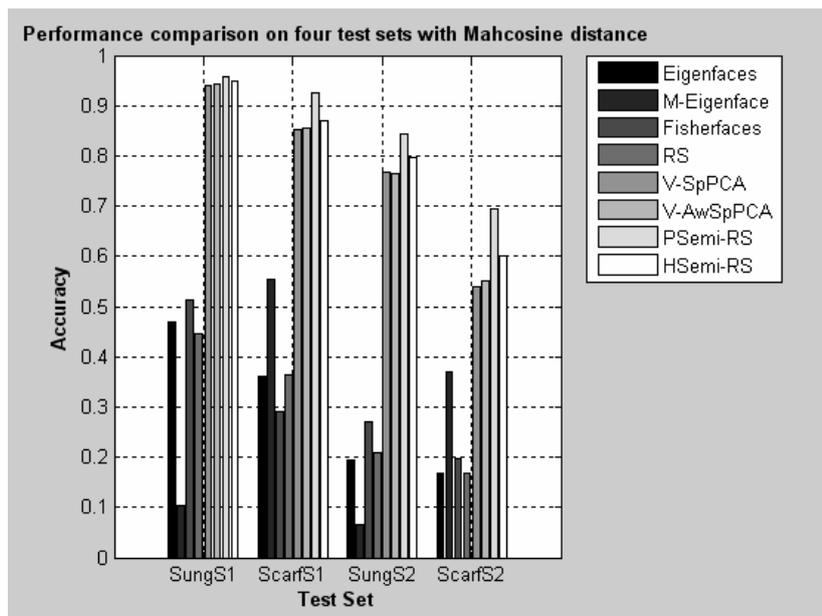
	Eigenfaces	M-Eigenfaces	Fisherfaces	Nitesh'RS	V-SpPCA	Aw-SpPCA	PSemi-RS	HSemi-RS
AR73Exp(M)	74.33	88.67	87.33	75.00	96.33	96.67	98.44	96.75
AR73Illu(M)	80.67	94.00	87.00	80.67	99.33	99.33	100	100
AR73Exp (E)	78.00	83.00	81.00	82.73	94.67	93.33	96.60	95.58
AR73Illu (E)	76.67	86.33	83.67	82.87	93.33	91.33	98.87	96.25

4.2.2 Occlusion on eyes and mouth

In this experiment, we aim to discuss the performance of Semi-RS when the test images are occluded by sunglass and scarf. In testing, four test sets in both session1 and session2 are selected. As illustrated in Fig. 5, both PSemi-RS and HSemi-RS outperform the other compared methods when all using the two distance metrics. Especially for Euclidean metric, both Semi-RSs' accuracies are respectively increased by at least 18% and 10%. For test set AR73SungS1, PSemi-RS can obtain over 96% recognition accuracy, while for the AR73Scarfs2 which is most difficult to be recognized in this database, PSemi-RS can still acquire recognition accuracy close to 70% using Mahcosine distance metric. These results indicate that our Semi-RSs are relatively robust to severe occlusions.



(a)



(b)

Fig.5. Performance comparison of methods on four test sets. (a), (b) compares the recognition accuracy on four test sets using Euclidean and Mahcosine metrics, respectively

4.3 Experiments on the Extended YALE Database B

The Extended YALE face database B contains more than 20000 single light source images of 38 subjects with 576 viewing conditions (9 poses \times 64 illumination conditions). In this paper, we use a subset provided and preprocessed by Lee [17]. This subset only contains those images with the frontal pose for

each individual, and includes 1920 image from 30 subjects (We only select the 1st-10th and 19th-38th individuals because there are some bad or damaged images in 11th -18th individuals in the provided image sets [17]). The original size 192×168 is resized to 48×42 and the same strategy as in [17] is adopted to divide the images of each individual into 5 subsets according to the angle of the light source direction. The images in SubSet1 are selected for training, and the images in the other subsets are selected for testing, respectively. All the examples of one person are shown in Fig.6.



Fig.6. All the samples of one person in Extended YALE B face database

Table 4. Results (%) on Extended Yale B face database with Euclidean metric

(The best performance in each case is bolded)

Euclidean Metric	Eigenfaces	M-Eigenfaces	Fisherfaces	Nitेश'RS	V-SpPCA	Aw-SpPCA	PSemi-RS	HSemi-RS
SubSet2	89.44	93.33	100	98.83	98.33	98.06	99.69	99.56
SubSet3	38.61	70.83	93.00	80.89	64.17	65.28	94.50	85.22
SubSet4	6.43	13.10	28.81	22.33	8.57	8.10	36.60	23.12
SubSet5	3.68	5.44	5.96	6.11	3.68	3.16	6.54	5.41

Table 5. Results (%) on Extended Yale B face database with Mahcosine metric

(The best performance in each case is bolded)

Mahcosine Metric	Eigenfaces	M-Eigenfaces	Fisherfaces	Nitesh'RS	V-SpPCA	Aw-SpPCA	PSemi-RS	HSemi-RS
SubSet2	100	94.72	100	93.33	98.89	99.17	99.67	99.56
SubSet3	99.73	87.11	99.17	86.39	100	100	100	100
SubSet4	74.52	26.19	45.00	43.57	89.05	89.76	94.98	90.71
SubSet5	27.54	5.96	10.35	18.25	82.98	83.16	89.56	85.41

Table 4 and Table 5 list the recognition results of Eigenfaces, M-Eigenfaces, Fisherfaces, Nitesh'RS, V-SpPCA, Aw-SpPCA and our Semi-RSs on four testing sets using Euclidean and Mahcosine distance metrics respectively. Comparing the results using Mahcosine with using Euclidean, we can draw a conclusion that, on the whole, Mahcosine distance metric is more suitable for recognizing images with severe illumination than Euclidean metric, so according to this, in the following analysis, we only discuss the results using Mahcosine distance. From Table 5, we can observe that 1) when the azimuth is very small, all the methods can gain high recognition accuracy, especially Eigenfaces and Fisherfaces methods, which achieve 100% performance on SubSet2. In this situation, the methods based on sub-image have no advantage over the holistic feature extraction methods. But with the increasing of lighting azimuth, the holistic feature extraction methods (Eigenfaces, Fisherfaces and Nitesh'RS) gradually present their non-robustness to lighting, while the sub-image-based methods have little effect. So properly partitioning an image into several sub-images is effective in improving the robustness of face recognition methods; 2) compared to the other methods, PSemi-RS gains the best recognition accuracy and at the same time, it shows PSemi-RS has better tolerance to illumination change. With the increasing of lighting azimuth, the maximal dropping difference in accuracy of PSemi-RS and HSemi-RS are 10.5% (from 100% to 89.56%) and 14.6% (from 100% to 85.41%), respectively, while the corresponding maximal dropping differences of Eigenface, M-Eigenface, Fisheface, Nitesh'RS, V-SpPCA and Aw-SpPCA are 72%, 88%, 89%, 75%, 16% and 16%, respectively, which indicates Semi-RSs are more robust to severe lighting.

Making a further comparison between the results in Table 4 and Table 5, we can find that the effectiveness of both PSemi-RS and HSemi-RS is quite insensitive to the used distance metrics, while the other methods are not. For example, Nitesh'RS is superior to V-SpPCA under Euclidean metric; while V-SpPCA is conversely superior to Nitesh'RS under Mahcosine metric. Our Semi-RSs yield consistent results on the two metrics and thus are more competitive and applicable than other compared methods.

4.4 Experiments on the FERET Database

This database is used to evaluate the recognition performance of Semi-RS in the scenario of only one training sample per person (in this case, Fisherfaces fails to work while Aw-SpPCA reduces to V-SpPCA). The selected FERET database contains 400 face images from 200 persons (71 females and 129 males), with the size of 256x384. Each person has two images (fa and fb) with different race, age, expression, illumination, occlusion, etc. In this experiment, the fa images are used for training and the fb images for testing, and all the faces are normalized (the line between the two eyes is parallel to the horizontal axis and the distance between the two eyes is set to 28 pixels) to 60x60 pixels.

We test the performance of our Semi-RSs based on both Euclidean and Mahcosine metrics. The results are presented in Table 6 from which we can find that PSemi-RS gains the best performance among all used methods including SVD perturbation which is especially designed for the recognition problem of one training sample per person. Although Hsemi-RS is slightly inferior to SVD perturbation under Mahcosine metric, wholly speaking, our proposed Semi-RSs are still effective in the scenario of one training image per person.

Table 6. Results (%) on FERET face database (The best performance in each case is bolded)

	Eigenfaces	M-Eigenfaces	Nitesh'RS	SVD perturbation	V-SpPCA	PSemi-RS	HSemi-RS
Accuracy(E)	84.50	82.00	84.60	86.00	84.50	88.80	86.25
Accuracy(M)	73.00	61.5	69.00	75	68.00	76.5	74.50

4.5 Experiments on the Yale Database

The Yale face database consists of 165 faces images of 15 individuals, each providing 11 image different images. The images are in upright, frontal position under various facial expressions and lighting conditions. In this experiment each image is manually cropped and resized to 32x32 pixels. A random subset with 5 images per person is taken to form training set and the rest of database is considered to be the testing set. The experiments are repeated 20 times for each method and the corresponding average accuracy and the standard deviation respectively are reported in Table 7. From the Table we can observe that our proposed methods not only gain higher accuracy but also have relatively smaller standard deviation, which means our methods are more stable.

Table 7. Average and standard deviation (%) on Yale face database
(The best performance in each case is bolded)

	Eigenface	M-Eigenface	Fisherfaces	Nitesh'RS	V-SpPCA	Aw-SpPCA	PSemi-RS	HSemi-RS
Accuracy(M)	52.9±4.9	69.33±4.39	68.4±4.3	59.4±4.4	68.8±3.1	68.6±4.2	74.5±3.4	71.8±2.7
Accuracy(E)	56.3±4.4	67.78±3.06	69.1±4.1	59.6±4.6	68.9±3.2	68.6±2.8	71.0±2.7	69.3±2.4

4.6 Experiments on ORL database

The ORL database contains the images from 40 subjects, with 10 different images for each subject. For some subjects, the images were taken at different sessions. There are variations in facial expressions (open or closed eyes, smiling or non-smiling), facial details (glasses or no glasses) and scale (up to about 10 percent). Moreover, the images were taken with a tolerance for tilting and rotation of the face of up to 20 degrees. All images are resized with a resolution of 32×32 pixels. Fig. 7 shows all the sample images of one person from the ORL database. In experiments, five images of each person are randomly selected for the training and the rest five for the testing. The averaged results of 20 runs of the experiments are reported in Table 8. The results reveal that Semi-RSs exhibit better performance to slight variations in pose angle and alignment. But it is worth to note that when there is more pose variation, these local feature-based methods including compared ones maybe fail to work well because local feature extracted from each local region is less helpful for correct recognition in this case.



Fig.7. Sample images from ORL database

Table 8. Results (%) on ORL face database (The best performance in each case is bolded)

	Eigenface	M-Eigenface	Fisherface	Nitesh'RS	V-SpPCA	Aw-SpPCA	PSemi-RS	HSemi-RS
Accuracy(M)	83.07	75.37	91.95	81.78	91.97	93.35	95.00	94.22
Accuracy(E)	87.80	85.75	94.30	91.70	91.90	93.10	94.50	93.35

4.7 Extended experiments based on LDA on AR database

In Section 3.2, we have mentioned that our Semi-RSs can be easily extended to other current subspace methods such as LDA [23] and NMF [25] etc. In order to verify such extensibility and the strength of combination between random subspace and local features, we use LDA as a substitute for PCA (named SemiRS-LDA) and conduct a preliminary experiment on AR database. The results are listed in Table 9 from which we can observe that both PSemiRS-LDA and HSemiRS_LDA are superior to all the

compared methods including the existing RSM methods RDA and R-LDA. Based on these preliminary results, we should have reason to believe that 1) Semi-RS has good extensibility to other current subspace methods; 2) combining random subspace and local features can effectively help to improve the classification performance in face recognition.

Table 9. Results (%) of LDA on AR face database using Euclidean and Mahcosine metrics

(The best performance in each case is bolded)

	M_Eigenface	Fisherfaces	Aw-SpPCA	RDA	R-LDA	PSemiRS-LDA	HSemiRS-LDA
AR73Exp(M)	88.67	87.33	96.67	85.62	88.00	98.27	97.17
AR73Illu(M)	94.00	87.00	99.33	89.81	92.00	100	99.63
AR73Sung1(M)	10.33	51.33	94.26	51.24	71.00	96.23	95.33
AR73Scarf1(M)	55.33	29.00	85.64	49.19	61.67	90.47	88.33
AR73Sung2(M)	6.67	27.00	76.40	28.05	34.00	81.57	78.90
AR73Scarf2(M)	37.00	19.67	55.00	22.67	29.33	63.07	55.33
AR73Exp(E)	83.00	81.00	93.33	86.50	87.13	98.80	96.73
AR73Illu(E)	86.33	83.67	91.33	88.79	90.20	99.73	99.00
AR73Sung1(E)	8.33	46.33	60.00	50.75	67.37	96.87	95.33
AR73Scarf1(E)	13.00	28.67	24.53	50.12	53.87	83.20	81.43
AR73Sung2(E)	6.00	24.00	33.00	27.88	36.93	82.27	75.33
AR73Scarf2(E)	7.33	16.33	11.35	23.21	23.60	50.73	48.17

4.8 Selection of Parameters

In order to study the influence of parameters (size of sub-image and random sampling rate) on the performance of Semi-RS, we conduct experiments by employing the parallel ensemble structure (PSemi-RS) and Euclidean distance metric on AR77, ARExp73 and ARIllu73 test sets.

The selection of random sampling rate (or the dimension of random sampled features) is an open problem for RSM based methods. Here, we will simply discuss the influence of random rate on performance of Semi-RS. Intuitively, the smaller the random rate is, the faster the algorithm performs, but at the same time, the more possibilities of missing informative features or missing dependences between features show. One extreme case is that only one pixel is selected as sampled feature from the feature set. In this situation, every sampled feature is independent and no structure information can be extracted, which is not desirable, so the random rate should not be too small. On the contrary, the random rate should also not be too big. The bigger the random rate is, the less the diversity between base classifiers is. Another extreme case is that the Semi-RS will reduce to V-SpPCA, when the random rate is 100%. So we should choose a proper random rate for Semi-RS. We set the size of sub-image to 6×6, and report the recognition

accuracies by varying random rate from 10% to 100% by interval 10% as in Fig.8 (a), from which we can observe that on the three testing sets, Semi-RS 1) shows similar change trends against the varying random rate; 2) can gain a relatively stable and high recognition performance when the random rate is between 30% and 50%.

Another parameter which needs to be selected in Semi-RS is the size of sub-image, it is an open issue as well. Similar to the above analysis, we can find that the smaller the size of sub-images is, the less the random sampled features is. When the size of sub-image is 1×1 , the sub-image reduces to a single pixel. On the contrary, the bigger the sub-image size is, the less local features the Semi-RS extracts. When the size of sub-images is equal to the size of original image, the Semi-RS reduces to the random sampling in ref [10] and little local information can be extracted. In order to examine effect of the sub-image size on the performance Semi-RS, we fix a random rate to 50% and take all possible sizes of sub-image (48 different partition-sizes in total). The experimental results show that PSemi-RS can gain over 95% recognition accuracy on the three testing sets when the sub-image size is taken as 6×3 , 6×4 , 6×6 , 6×8 , 11×6 , 11×8 and 11×12 , respectively, while for the rest sub-image sizes, the accuracies are consistently below 95%. As a result, we only report these results with respect to those representative sub-image sizes listed in $\{2 \times 2, 3 \times 3, 6 \times 3, 6 \times 4, 6 \times 6, 6 \times 8, 11 \times 6, 11 \times 8, 11 \times 12, 11 \times 24, 22 \times 4, 22 \times 16, 33 \times 8, 33 \times 16\}$ in Fig.8 (b). These results verify our intuitional analysis, i.e., the size of sub-image should be neither too small nor too big, but how to choose a proper size of sub-image is problem-dependent. Fortunately, we find an effective method to avoid the selection of sub-image size. Adopting the strategy as in [7-9], we randomly select the size of sub-image and construct multiple classifiers with this sub-image size, and then combine all the classifiers with parallel combination structure. The motivation behind this strategy is to adopt a combination approach of classifiers associated to different parameters to avoid selecting a single set of optimal parameters for a classifier. Table 10 demonstrates that the performance of such multiple-parameters ensemble is close or slightly inferior to that of PSemi-RS with optimal recognition performance.

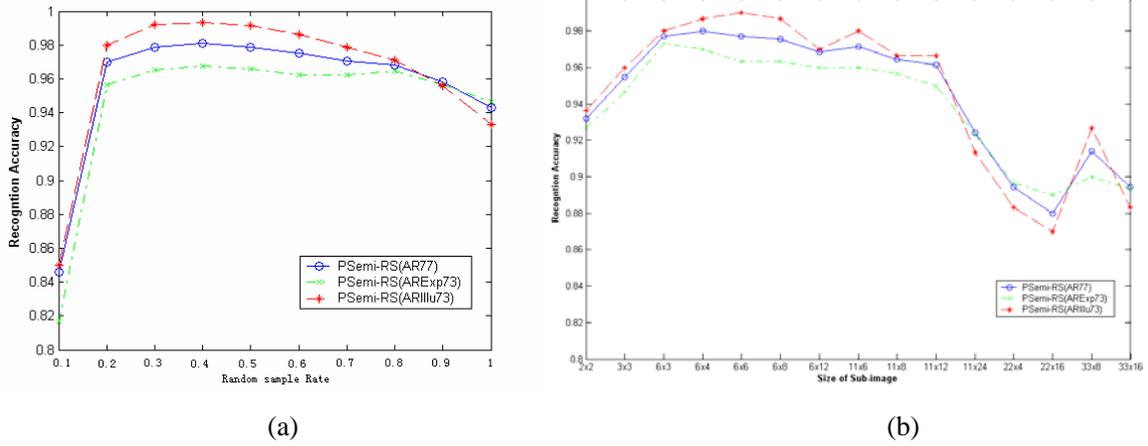


Fig.8. Influence of Parameters on performance

Table 10. Performance comparison between ensemble of parameters and PSemi-RS (%)

	AR77	ARExp73	ARIllu73
multiple-parameters ensemble	98.00	96.67	99.00
PSemi-RS (max)	98.00	97.33	99.00

5. Diversity-Error Analysis

Diversity between classifiers is a key property of classifier ensemble. High diversity assures that different base classifiers make different error on different patterns, which means that, by combining classifiers, one can arrive at an ensemble with more accurate decisions. So it greatly affects recognition performance of classifier ensemble. Tumer has pointed out that the ensemble error is in inverse proportion to the diversity between classifiers [28]. Hence in order to improve the recognition performance of ensemble, it is reasonable to create classifiers with more diversity. Here, we will adopt Kappa diversity-error diagram [29] to analyze the diversity of proposed Semi-RS, Nitesh'RS [10] and V-SpPCA.

Kappa diversity-error diagram [29] is an effective pair-wise diversity measure method which evaluates the level of agreement (or diversity) between two classifiers and averaged error of the pair of classifiers. To get the agreement between two classifiers C_i and C_j , a coincidence matrix M of the two classifiers should be first constructed. Let there be N patterns for testing and the entry $m_{k,s}$ of M denotes the number of x for which $C_i(x) = k$ and $C_j(x) = s$. The agreement [29] between C_i and C_j can be defined as follows:

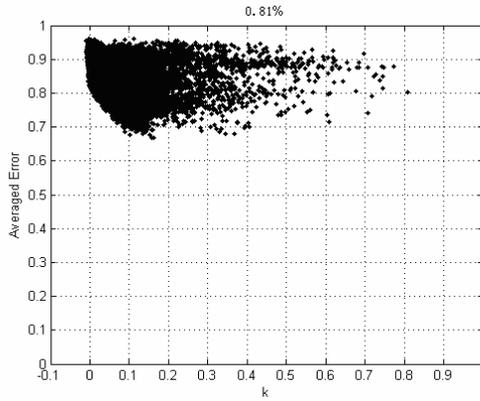
$$\kappa_{i,j} = \frac{\sum_t \frac{m_{tt}}{N} - ABC}{1 - ABC} \quad (4)$$

Where $\sum_t \frac{m_{tt}}{N}$ is the probability that the two classifiers agree and ABC is "agreement-by-chance"

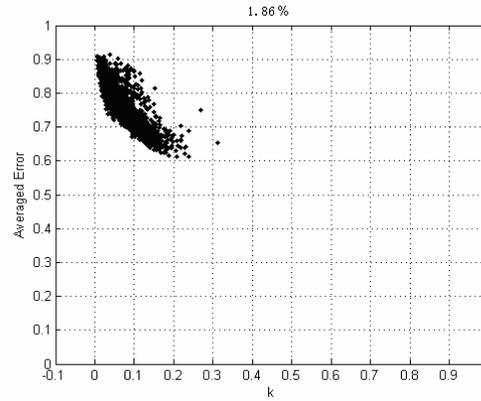
$$ABC = \sum_u \left(\sum_v \frac{m_{u,v}}{N} \right) \left(\sum_v \frac{m_{v,u}}{N} \right) \quad (5)$$

$\kappa=0$ when the agreement of the two classifiers equals that expected by chance, $\kappa=1$ when the two classifiers produce identical class labels for all test pattern, and a more desirable case happens when $\kappa < 0$, in this condition, agreement is weaker than expected by chance.

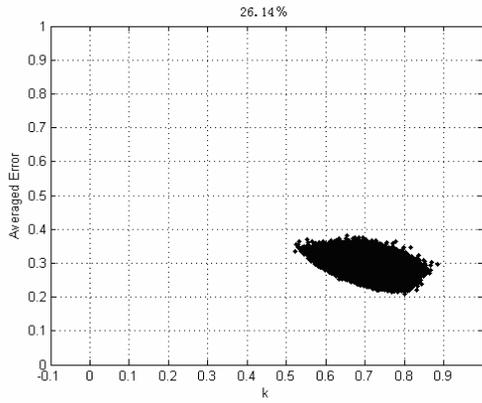
Fig.9 and Fig.10 illustrate the kappa-error diagrams of different methods for AR and EYALE datasets, respectively. In these diagrams, the x -axis is the agreement (κ) between pairs and the y -axis is the averaged individual error (E) of pair classifiers. As small values of κ indicate better diversity and small values of E indicate better accuracy, the most desirable pairs of classifiers will lie in the bottom left corner. In experiments, in order to compare the diversity of different methods under the same condition, we set the random sampling rate to 50% for both Nitesh'RS and Semi-RS, and select two different ensemble sizes for Nitesh'RS: one is equal to the ensemble size of V-SpPCA (the ensemble size of V-SpPCA is decided by the number of sub-images) and the other is the ensemble size of Semi-RS. The base classifier is the nearest neighbor classifier with Mahcosine distance metric. For the AR dataset, the ensembles are trained on the first seven face images of session1 and the kappa-error diagram are calculated on the first seven face images of session2. While for the EYALE dataset, the ensembles are performed on the SubSet1 and the kappa-error diagram are evaluated on the joint set of SubSet2 and SubSet3.



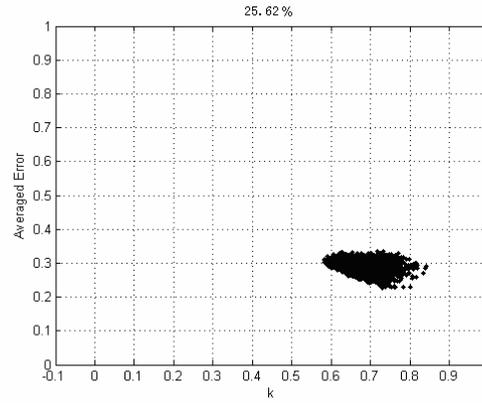
(a)



(b)

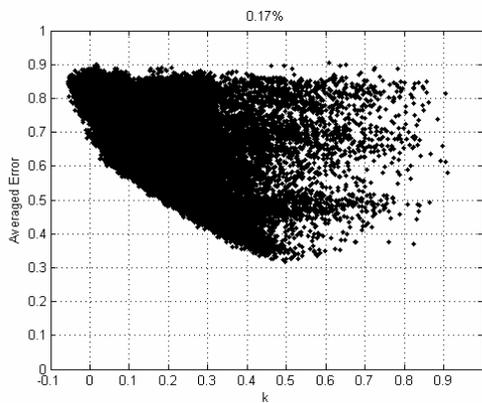


(c)

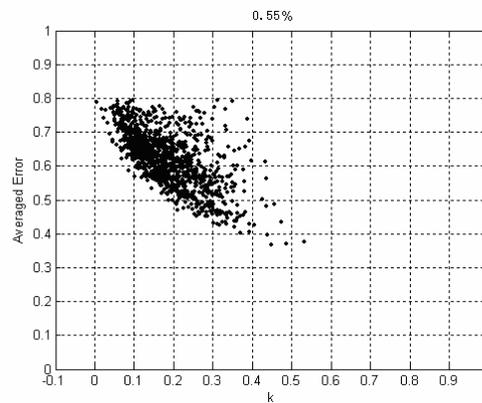


(d)

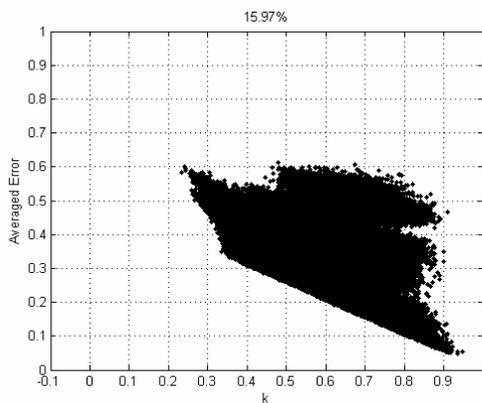
Fig.9 Kappa-Error diagrams on AR database (a) Semi-RS, (b) V-SpPCA, (c) Nitesh'RS1760, (d) Nitesh'RS88. The ensemble error on the testing set is displayed above the plot.



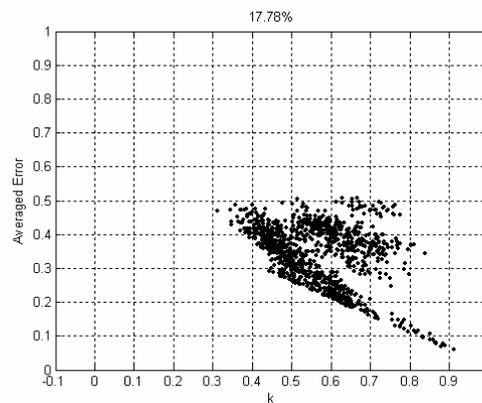
(a)



(b)



(c)



(d)

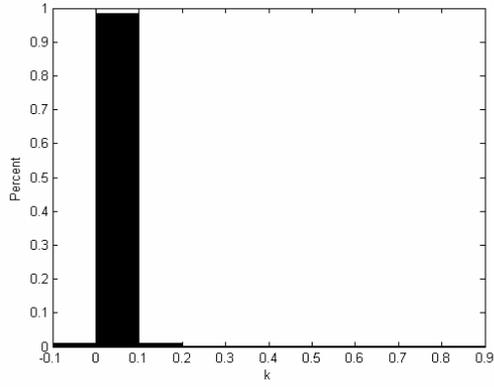
Fig.10. Kappa-Error diagrams on EYALE database. (a) Semi-RS, (b) V-SpPCA, (c) Nitesh'RS960, (d) Nitesh'RS48. The ensemble error on the testing set is displayed above the plot.

We compare the diversity of different methods from the following aspects: 1) comparing (a) and (c), which are the kappa-error diagrams of proposed Semi-RS and of Nitesh'RS [10] respectively and they are have the same ensemble size, we can observe that Nitesh'RS has the lower averaged error but the smaller

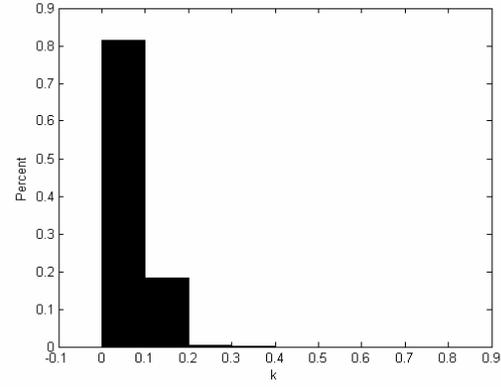
diversity than Semi-RS. Since mass of pair points lie in the high agreement range, the ensemble performance of Nitesh'RS is worse than that of Semi-RS; 2) comparing the kappa-error diagrams of Nitesh'RS with different ensemble sizes ((c) and (d)), although (c) uses more classifiers, increasing the number of classifiers does not provide any sizeable gains. The main reason is that the base classifiers constructed in Nitesh'RS have less diversity thus it is not quite easy to make different errors on different patterns; 3) comparing the diversity of V-SpPCA and Nitesh'RS ((b) and (d)), which use the same ensemble size, it is obviously observed that V-SpPCA is more diverse than Nitesh'RS; 4) comparing the (a) and (b), it seems that from the figures, the diversity of Semi-RS is worse than that of V-SpPCA since the Semi-RS spans larger range than V-SpPCA does and mass of pair points lie in the high agreement range, but the fact is not true. In order to explain this, we make a further analysis for the kappa-errors diagrams. Fig.10 shows the proportions of all the pair points of V-SpPCA and those of Semi-RS respectively under different diversity range. From Fig. 10, we can observe that compared with V-SpPCA, Semi-RS has relatively larger proportion of the pair points which centralize the larger diversity range. More specifically, for AR database, about 98% of all pair points lie in the agreement range $[0, 0.1]$ for Semi-RS, while 81% for V-SpPCA. This indicates that Semi-RS has, on the whole, much larger diversity than V-SpPCA. For EYALE, we also draw the similar outcomes. We make a further explanation for Fig.10 from another angle. For this, let us define a new measure within certain given range $[a, b]$ of κ , called the κ -weighted area here, as follows:

$$\kappa WA = 100 \times \int_a^b (\kappa \times Percent_{\kappa}) d\kappa \quad (6)$$

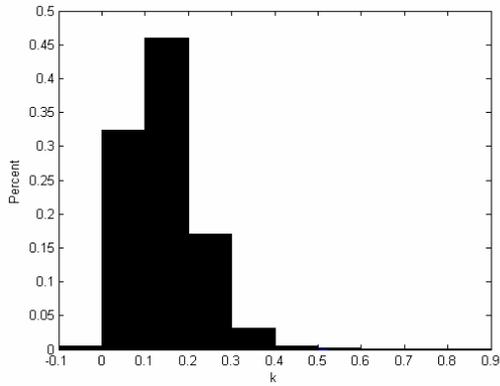
where $Percent_{\kappa}$ denotes the corresponding value of y -axis when the value of x -axis is κ . Since there exists negative κ , κWA can be negative. In fact, the κWA characterizes the whole diversity of ensemble method. The less the κWA is, the larger the diversity is. Table 11 lists the κWA s under the whole diversity range corresponding to Semi-RS and V-SpPCA. From the Table, Semi-RS has less κWA than V-SpPCA, thus indicating the former has larger diversity than the latter. In addition, it is much more worthwhile emphasizing that in Semi-RS, there exist many negative values of κ , according to [30], the more the number of the negative value, the better the diversity, as a result, stating that Semi-RS can have higher diversity than V-SpPCA.



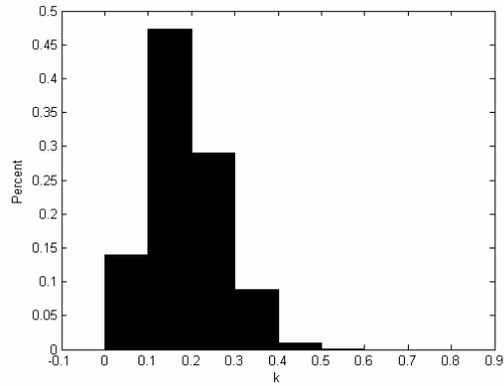
(a)



(b)



(c)



(d)

Fig.11. the proportion of pair points under different diversity ranges (a) Semi-RS for AR, (b) V-SpPCA for AR, (c) Semi-RS for EYALE and (d) V-SpPCA for EYALE

Table 11. the κ WA of Semi-RS and V-SpPCA on AR and EYALE database

	Semi-RS (AR)	V-SpPCA(AR)	Semi-RS (EYALE)	V-SpPCA (EYALE)
κ WA	0.5047	0.6916	1.3892	1.8509

A problem worthy to be discussed is the trade-off between diversity and accuracy of base classifier which is also an importance factor of classifier ensemble. For a good ensemble method, the accuracy of its base classifier must be better than random guess and the classifiers must be diverse. So we always expect that the base classifiers should be as accurate and diverse as possible. However, both diversity and accuracy are inversely related, thus generally, it is impossible for the base classifiers to be diverse if the base classifiers have high accuracies, and vice versa. This leads to the famous accuracy-diversity dilemma. The kappa-error diagrams shown in Fig.9 and Fig.10 indicate this dilemma: Semi-RS and V-SpPCA have larger diversities than Nitesh'RS, but inversely, they have lower averaged accuracies than the latter. So it is expected that there will be a trade-off between diversity and accuracy to improve the performance of

ensemble. Unfortunately, until now, no general method has been proposed to show how to compromise between accuracy and diversity of base classifier.

6. Conclusion

In this paper, we propose a Semi-RS framework especially for face recognition. Unlike traditional random subspace method that completely randomly selects the feature from the whole sample feature set, Semi-RS considers the distribution of face features and makes use of the local spatial relation between pixels of face image. It applies random sampling features to each local region (or a sub-image) partitioned from the original face image. As a result, Semi-RS fuses the advantages of RSM and sub-image method so that it can effectively solve the small sample size problem and improves robustness to such variations as illumination, occlusion, etc. We take PCA as an example to formulate the Semi-RS and evaluate its performance on five benchmark face image databases using two distance metrics, and the experimental results indicate that Semi-RS is more robust, practical and competitive than both the existing sub-image methods and the random sampling methods. In addition, we also discuss the diversities of different methods using kappa-error diagram, and the results show that Semi-RS has higher diversity than both RSM and V-SpPCA. In our future work, we will try to embed the idea of Semi-RS into traditional methods such as CCA [24], NMF [25], etc. Moreover, we will try to exploit more complex combination rule to combine base classifiers.

Acknowledgment

This work was supported by National Natural Science Foundations of China (60773060 and 60773061) and Natural Science Foundations of Jiangsu Province (BK2006187 and BK2008xxx). We also thank Prof. Xiaoyang Tan for beneficial discussion.

Reference:

- [1] W. Yu, X. Teng and C. Liu, Face recognition using discriminant locality preserving projections, *Image and Vision Computing*, 24(3) (2006), pp 239-248.
- [2] A. Srivastava, X. Liu, Face recognition using optimal linear bases of range images, *Image and Vision Computing*, 24(3) (2006), pp 291-299.
- [3] X. Y. Tan, S.C. Chen, Z.H. Zhou, and F. Zhang, Face recognition from a single image per person: A

survey, *Pattern Recognition*, 39(9) (2006), pp 1725–1745.

[4] M. Turk and A. Pentland, Eigenfaces for recognition, *Journal of Cognitive Neuroscience*, 3(1) (1991), pp 71–96.

[5] 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) (1997), pp 711–720.

[6] T. K Ho, The random subspace method for constructing decision forests, *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 20(8) (1998), pp 832-844.

[7] X. G. Wang, X.O.Tang, Random Sampling for Subspace Face Recognition, *International Journal of Computer Vision*, 70(11) (2006), pp 91-104

[8] X. G. Wang, X.O. Tang, Random sampling LDA for face recognition. *Proceedings of Computer Vision and Pattern Recognition*, vol. 2 (2004), pp 259–265.

[9] X. G. Wang, X.O. Tang, Subspace analysis using random mixture models, *IEEE Conference on Computer Vision and Pattern Recognition*, 2005, pp 574-580.

[10] V.Nitesh. Chawla and KevinW. Bowyer, Random subspaces and subsampling for 2-D face recognition, *IEEE Conference on Computer Vision and Pattern Recognition*, (2005) pp. 582-589.

[11] X. Zhang, Y. Jia, A linear discriminant analysis framework based on random subspace for face recognition, *Pattern Recognition*, 40(9) (2007), pp 2585-2591.

[12] H. Yu, J. Yang, A direct LDA algorithm for high-dimensional data-with application to face recognition, *Pattern Recognition*, 34(10) (2001), pp 2067–2070.

[13] R. Gottumukkal, V. K. Asari, An improved face recognition technique based on modular PCA approach, *Pattern Recognition Letters*, 25(4) (2004), pp 429-436

[14] S. C. Chen and Y. L. Zhu, Subpattern-based principle base analysis, *Pattern Recognition*, 37(1) (2004), pp 1081-1083

[15] K. R. Tan and S. C. Chen, Adaptively weighted sub-pattern PCA for face recognition, *Neurocomputing*, vol 64 (2005), pp 505-511

[16] A.M. Martinez, R. Benavente, The AR face database. CVC Technical Report #24, June 1998

[17] 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) (2005), pp 1–15

[18] D.C.Tao and X.O.Tang, Asymmetric bagging and random subspace for support vector machines-based

relevance feedback in image retrieval, *IEEE Transactions on Pattern Analysis and Machine Intelligence* 28(7) (2006), pp 1088-1098.

[19] P. J. Phillips, H. Wechsler, J. Huang, P. J. Rauss, The FERET database and evaluation procedure for face recognition algorithms, *Image and Vision Computing*, 16(5) (1998), pp 295-306.

[20] 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) (2005), pp 128– 138.

[21] A. Pentland, B. Moghaddam, and T. Starner. View-based and Modular Eigenspaces for Face Recognition. *Proceedings of Computer Vision and Pattern Recognition*, (1994), pp 84-91.

[22] D.Q.Zhang, S.C.Chen, and Z.-H.Zhou, A new face recognition method based on SVD perturbation for single example image per person, *Applied Mathematics and Computation*, 163(2) (2005), pp 895-907.

[23] K.Fukunage, *Introduction to statistical pattern recognition*, second edition (1990), Academic Press, Boston.MA.

[24] D.R. Hardoon, S. Szedmak and J. Shawe-Taylor, Canonical correlation analysis: an overview with application to learning methods, *Neural Computation*, 16 (2004), pp 2639-2664.

[25] D.D.Lee and H.S.Seung, Algorithms for non-negative matrix factorization, *Neural Information Processing Systems*, 13 (2001), pp 556–562.

[26] <http://cvc.yale.edu/projects/yalefaces/yalefaces.html>

[27] <http://www.cl.cam.ac.uk/Research/DTG/attarchive/facedatabase.html>

[28] K. Tumer, N.C. Oza, Error correlation and error reduction in ensemble classifiers, *Connection science*, 8 (3&4) (1996), pp 384-404.

[29] D.D.Margineantu and T.G.Dietterich, Pruning adaptive boosting, *International Conference of Machine Learning* (1997), pp 211-218.

[30] J.J. Rodriguez, L.I. Kuncheva, Rotation forest: a new classifier ensemble method, *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 28(10) (2006), pp 1619-1630.

[31] Z. H. Zhou and Y. Yang, Ensembling local learners through multimodal perturbation, *IEEE Transactions on Systems, Man, and Cybernetics-Part B*: 35(4) (2005), pp 725-735

[32] G. D. Su, C. P. Zhang, R. Ding and C. Du, MMP-PCA face recognition method, *ELECTRONICS LETTERS*, 38(25) (2002), pp 1564-1566