Recognizing Partially Occluded, Expression Variant Faces from Single Training Image per Person with SOM and soft *k*NN Ensemble

Xiaoyang Tan, Songcan Chen, Zhi-Hua Zhou, *Member, IEEE*, and Fuyan Zhang

***Abstract*—Most classical template-based frontal face recognition techniques assume that multiple images per person are available for training, while in many real-world applications only one training image per person is available and the test images may be partially occluded or may vary in expressions. This paper addresses those problems by extending a previous local probabilistic approach presented by Martinez, using the Self-Organizing Map (SOM) instead of a mixture of Gaussians to learn the subspace that represented each individual. Based on the localization of the training images, two strategies of learning the SOM topological space are proposed, namely to train a single SOM map for all the samples and to train a separate SOM map for each class, respectively. A soft *k* nearest neighbor (soft *k*-NN) ensemble method, which can effectively exploit the outputs of the SOM topological space, is also proposed to identify the unlabelled subjects. Experiments show that the proposed method exhibits high robust performance against the partial occlusions and variant expressions.**

***Index Terms*—Face recognition, single training image per person, occlusion, face expression, self-organizing map**

I. INTRODUCTION

As one of the few biometric methods that possess the merits of both high accuracy and low intrusiveness, Face Recognition Technology (FRT) has a variety of potential applications in information security, law enforcement and surveillance, smart cards, access control, among others [1-3]. For this reason, FRT has received significantly increased attention from both the academic and industrial communities during the past twenty years. Numerous recognition methods have been proposed, some of which have obtained much success under constrained conditions [3]. However, the general face recognition problem is still unsolved due to its inherent complexity.

The complexities of face recognition mainly lie in the constantly changing appearance of human face, such as variations in occlusion, illumination and expression. One way to overcome these difficulties is to explain the variations by explicitly modeling them as free parameters [4-5]. On the other hand, the variations can be attacked indirectly by searching one or more face subspaces of the face so as to lower the influence of the variations. The early way to construct face subspaces is by manually measuring the geometric configural features of the face [8]. Later, researchers realize that one can extract not only the texture and shape information but also the configural information of the face directly from its raw-pixels-based representation [9]. In addition, a higher recognition rate can also be achieved with latter method as compared to the geometric-based approach [8]. Consequently, template-based methods have become one of the dominant techniques in the field of face recognition since the 1990s.

To construct the face subspaces with good generalization, a large and representative training data set should be required due to the high dimensions of the face images [10]. However, this is not always possible in many real world tasks, such as finding a person within a large database of faces (e.g. in the law enforcement scenarios), typically only one image is available per person. This brings much trouble to many existing algorithms. Most subspace methods such as Linear Discriminant Analysis (LDA) [12-14], Bayesian matching methods [16] and Evolutionary Pursuit (EP) [35] may fail because the intra-class variation becomes zero when there is only one image per class for training. Furthermore, even the parameter estimation becomes possible when two or more samples are available, these algorithms' performance may still suffer a lot from the small, non-representative sample sets [17-18].

In fact, this so-called *one image per person* problem can be traced back to the early period when the geometric-based methods were popular, where various configural features such as the distance between two eyes are manually extracted from the single face image [8]. Recently, several researchers [18-20]

Manuscript received Jan. 12, 2004. The work was supported by National Natural Science Foundation of China under Grant No. 60271017, National Outstanding Youth Foundation of China under Grant No. 60325207, Jiangsu Science Foundation Key Project under Grant No. BK2004001, For Ying Tung Education Foundation under Grant No. 91067, and Excellent Young Teachers Program of MOE of China.

Xiaoyang Tan is with the National Laboratory for Novel Software Technology, Nanjing University, Nanjing 210093, and the Department of Computer Science and Engineering, Nanjing University of Aeronautics & Astronautics, Nanjing 210016, China. (e-mail: x.tan@nuaa.edu.cn).

Songcan Chen is with the Department of Computer Science and Engineering, Nanjing University of Aeronautics & Astronautics, Nanjing 210016, and the Shanghai Key Laboratory of Intelligent Information Processing at Fudan University, China. (e-mail: s.chen@nuaa.edu.cn).

Zhi-Hua Zhou is with the National Laboratory for Novel Software Technology, Nanjing University, Nanjing 210093, China. (e-mail: zhouzh@nju.edu.cn).

Fuyan Zhang is with the National Laboratory for Novel Software Technology, Nanjing University, Nanjing 210093, China. (e-mail: fyzhang@nju.edu.cn).

have begun to again pay attention to this classical problem within the template-based framework due to the needs of the applications and its potential significance of solving the likewise small sample problem.

The general idea behind these literatures to solve this problem is to try to squeeze as much information as possible from the single face image, which is used to provide each person with several imitated face images. For example, Chen et al. enlarged the training image database using a series of *n*-order projected images [19]. Beymer and Poggio developed a method to generate virtual views by exploiting prior knowledge of faces to deal with the pose-invariant problem [20]. Improved recognition accuracies have been achieved by these methods. However, as Martinez stated [18], one non-ignorable drawback of these methods is that it may be highly correlated among the generated virtual images and therefore these samples should not be considered as independent training images.

Besides the one image per person problem, there exist other problems that make things even more complicated, such as the *partial occlusion* and/or *expression-invariant* problem. The former means that portions of the face image are missed or occluded due to glasses or clothing, while the latter means that the facial expressions in the training images are different from those in the testing images of the same subject. Both the problems are supposed to be difficult for the traditional template-based paradigm, such as principal component analysis (PCA) [27-29].

Recently, Martinez [18, 21-22] has partially tackled the above problems using a local probabilistic method, where the subspace of each individual is learned and represented by a separate Gaussian distribution. This paper extends his work by proposing an alternative way of representing the face subspace with Self-Organizing Maps (SOM, [24-25]). One of the main motivations of such an extension is that, even when the sample size is too small to faithfully represent the underlying distribution (e.g. when not enough or even no virtual samples are generated), the SOM algorithm can still extract all the significant information of local facial features due to the algorithm's unsupervised and nonparametric characteristic, while eliminating possible faults like noise, outliers, or missing values. In this way, the compact and robust representation of the subspace can be reliably learned. Furthermore, a soft k nearest neighbor (soft k-NN) ensemble method, which can efficiently exploit the outputs of the SOM topological space, is also proposed to identify the unlabelled subjects.

The paper proceeds as follows. The local probabilistic method proposed by Martinez is briefly introduced in section 2. The proposed method is described in Section 3. The experiments are reported in section 4. Finally, conclusions are drawn in section 5.

II. Local Probabilistic Approach

The local probabilistic approach works as follows [18, 21]: firstly, a set of virtual images accounting for all possible localization errors of the original image are synthetically generated for each training face. And then, each face (including the generated face images) is divided into six local areas, and the subspace of every subimage is estimated by means of a mixture model of Gaussians using the EM algorithm. Finally, the eigen-representaton of each local areas are calcualted within their own subspace, and each sample image can thus be represented as a mixture of Gaussians in each of these eigenspaces.

In the identification stage, the test images are also divided into six local areas and are then projected onto the above computed eigenspaces. A probabilistic rather than a voting approach is used to measure the similarity of a given match. Experiments on a set of 2600 images show that the local probabilistic approach does not reduce accuracy when 1/6 of the face is occluded (on the average).

A weighted local probabilistic method is also proposed by the author to address the expression-invariant problem [18, 22]. The idea is based on the fact that different facial expressions influence different parts of the face more than others, thus a learning mechanism is proposed to learn such prior information by weighting each of the local parts with the values that vary depending on the facial expression displayed on the testing image.

The (weighted) local probabilistic method greatly improves the robustness of the recognition system. However, the mixture of Gaussians is a parametric method which heavily depends on the assumption that the underlying distribution is faithfully represented with a lot of samples. While those samples can be synthetically generated as the way described above, the computational and storage costs along with the procedure of generating virtual samples may be very high (e.g. 6,615 samples per individual in [18]) when the face database is very large. We extend the method using an unsupervised and nonparametric method, i.e. SOM, which can represent the subspace of local features reliably even no extra virtual samples are generated, and the application scope of the method is hereby expanded. The proposed method will be detailed in the next section.

III. The Proposed Method

A high-level block diagram of the proposed method is shown in Fig.1. The details of the method are described in the following subsections.

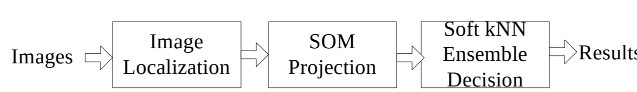


Fig.1. Block diagram of the proposed method for face recognition.

A. *Localizing the face image*

In the case of only limited training samples available per person, it is almost unavoidable to face the dilemma of high dimensions of image data and small samples. One way to deal with this difficulty is to reduce the dimensionality using projection methods such as principal component analysis (PCA) [27-29], however, one of the main drawbacks of the classical

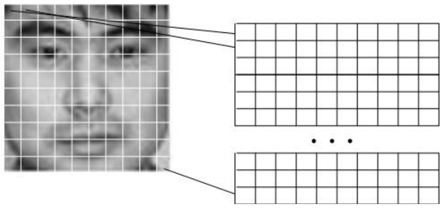


Fig. 2. Representation of a face image by a set of sub-block vectors

PCA projection is that it is easy to be subject to gross variations and thus sensitive to any changes in expression, illumination, etc. Most recently, Yang et al. have proposed a new technique named two-dimensional principal component analysis (2DPCA) [43], which is based on small 2D image matrices instead of 1D vectors and thus more suitable for small sample size problem than classical PCA.

An alterative way is to use local approaches [8, 11, 16, 23, 28, 30, 31,34,48], i.e., based on some partition of the image, the original face can be represented by several *low* dimensional local feature vectors (LFVs) rather than one full *high* dimensional vector, thus the small sample problem can be alleviated. Moreover, the sub-pattern dividing process can also help increase the diversity [31], making the common and class-specific local features be identified more easily [30]. These advantages are useful for the identification of the unknown persons.

In the implementation of this paper, the original image is divided into $M(=l/d)$ non-overlapping sub-blocks with equal size. The M LFVs are obtained by concatenating the pixels of each sub-block, where l and d are the dimensionalities of the whole image and each sub-block, respectively (Fig. 2.). Note that other local features can also be used, such as the local eigenfeatures [16, 21, 28] and local Gabor filters [36]. For example, one can extract local features at different scale as done in [50] to deal with the multi-scale problem.

B. The Use of Self-Organizing Map

1) The Single SOM-face Strategy

A large number of sub-blocks may be generated from the previous localizing stage. Now we need an efficient method to find the patterns or structures of the sub-blocks without any assumption about their distribution. There exist several classical methods for this purpose, such as the LGB [32], SOM [24,25], and fuzzy C-means [33]. Fleming et al.'s early work also suggested that unsupervised feature clustering helps to improve the performance of the recognition system [37].

The Self-Organizing Map (SOM), which approximates an unlimited number of input items by a finite set of weight vectors, is chosen here for several reasons as follows: Firstly, the SOM learning is efficient and effective, suitable for high dimensional process [25]. Secondly, it is found that the SOM algorithm is more robust to initialization than other algorithm such as LGB [38]. Finally and most importantly, in addition to clustering, the weight vectors can be organized in an ordered manner so that the topologically close neurons are sensitive to similar input sub-blocks [25].

In this paper, the original SOM algorithm is used, which is arguably one of the most computationally efficient [39]. This is very important, especially when the number of sub-blocks is very large. Furthermore, to accelerate the computation of the SOM, the batch formulation of the SOM algorithm is used, which has the advantages of fast convergence and being less sensitive to the order of presentation of the data. Recently, the batch-SOM algorithm has been employed by Kohonen et al. to fast learn a large text corpus in their WEBSOM project [39].

Formally, let $X(t)=\{x_i(t)\mid i=1,\cdots,N\}$ be the set of input vectors at time t, and $A=\{e_1,e_2,\cdots,e_Q\}$ be the neurons of the SOM respectively. Also let the weight vector (also called codebook or reference vectors) stored in the neuron $e_i(i\in[1,Q])$ be w_i, which in turn decides the location of the neuron e_i in the lattice A. The Voronoi set of the neuron e_i is denoted by V_i, which consists of all the n_i closest sub-blocks of the weight vector w_i in the input space. Let N denote the size of the data set and Q, the number of neurons in the lattice, respectively. After initialization, the batch-mode SOM algorithm consists of iterating the following three steps to adjust the weight vectors until they can be regarded as stationary [25, 40]:

Step1: Partition all the sub-blocks into Voronoi regions $V_i(i\in[1,Q])$ by finding each sub-block's closest weight vector according to: $c=\arg\min_i\{\|x_i(t)-w_i(t)\|\}$, where c is index of the winner neuron.

Setp2: And then the average of the sub-block vectors $x(t)$ over V_i, denoted by $\bar{x}_i$, is computed by: $\forall i,\bar{x}_i=\sum_{x(t)\in V_i}x(t)\Big/n_i$,with $N=\sum_i n_i$

Setp3: Smooth the weight vector of each neuron: $w_i^*=\sum_j n_j h_{jr}\bar{x}_j\Big/\sum_j n_j h_{jr}$, where $h_{j,r}$ is a neighborhood function which governs both the self-organization process and the topographic properties of the map.

After the SOM map has been trained, all the sub-blocks from each training face are mapped to the Best Matching Units (BMUs) in the SOM topological space by a nearest neighbor strategy. The corresponding weight vectors of the BMUs will be used as the "prototype" vectors of each class for later recognition purpose (detailed in next section). Fig.3 shows an example of an original image, its projection and the reconstructed image (called "SOM-face", constructed with the corresponding prototype vectors).

2) The Multiple SOM-face strategy

One drawback of the manifold-based learning methods is the requirement of the recomputation of the base vectors (such as eigenvectors in PCA, weight vectors in SOM, etc.) when new individuals are presented. One way to deal with this problem is to train a separate subspace for each face, that is, when a new face is encountered, only the new one rather than the original whole database is needed to be re-learned, thus the

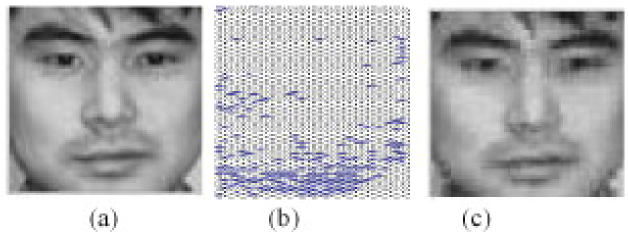


Fig.3. Example of an original image (a), its projection (b) and the reconstructed image(c)

computational cost can be significantly reduced due to the simplification of the learning procedure.

Based on the above idea, another SOM-face strategy, namely the multiple SOM-face (MSOM-face) strategy is proposed (similar ideas can be found in [15] and [26]). Formally, let the j^{th} new face image be presented to the system, which is divided into a set of equally-sized sub-blocks, denoted as $\left\{x_i^j \mid i=1,\cdots,M\right\}$. Then a separate small SOM map S^j for the face will be trained with its class-designated samples, using the previously-described batch-mode SOM algorithm. To evaluate the prototype vectors of the new class, we can map the new face's sub-blocks onto the newly trained map. Another way is to combine all the small maps into one big map and then recalculate all the prototype vectors for every class. We have found that the latter method is more accurate when a large number of classes exist.

C. Identify Faces based on SOM-face

To identify an unlabelled face, a classifier should be built on the SOM map. Popular supervised classifiers, such as LVQ, MLP, SVM, etc, require that the neurons of the SOM be labeled before being used for recognition. Unfortunately, labeling neuron may also lead to the losing of information. For example, in a commonly used labeling mechanism, majority voting (MV), each neuron is hard labeled according to the maximum class frequency obtained from the training data, while a large amount of information about the classes other than the winner class is thus lost from the neuron.

For this reason, a soft k nearest neighbor (kNN) ensemble decision scheme (Fig. 4) is proposed to avoid the above problem and to effectively exploit as much information as possible from the outputs of the SOM topological space. In the

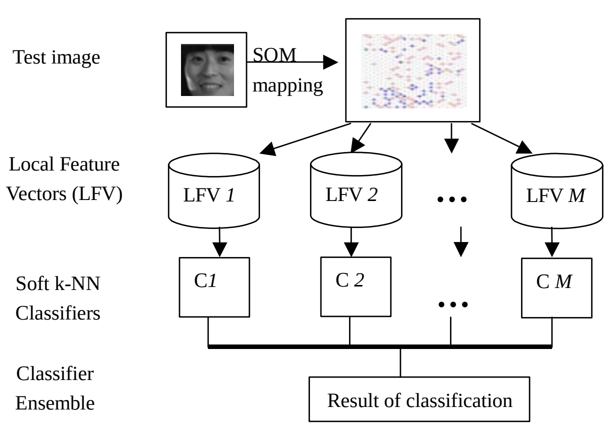


Fig. 4. The architecture of the soft kNN ensemble classifier

proposed method, each component classifier is supposed to output a confidence vector which gives the degree of support for each LFV's membership in every class, and then all the outputs will be combined with a sum aggregation method [49] to give the final decision. The details of the proposed method are described below.

Given C classes, to decide which class the test face x belongs to, we first divide the test face into M non-overlapping sub-blocks as mentioned before, and then project those sub-blocks onto the trained SOM maps to obtain the face's SOM-face representations.

Then, a distance matrix, describing the dissimilarity between the test face and every training face in the current SOM topological space, is calculated. The distance-calculation algorithm will be detailed in Table I. Its outputs, however, are introduced directly below for convenience of description, which take the form of a $M\times C$ matrix as follows:

$$D(x)=\begin{bmatrix} d_{11} & d_{12} & \cdots & d_{1C} \\ d_{21} & d_{22} & \cdots & d_{2C} \\ \cdots & \cdots & \cdots & \cdots \\ d_{M1} & d_{M2} & \cdots & d_{MC} \end{bmatrix} \equiv [dv_1 \quad \cdots \quad dv_M]^T \qquad (1)$$

$$dv_j = \left[d_{j1}, d_{j2}, \quad \ldots \quad , d_{jC}\right] \qquad (2)$$

where dv_j is called the distance vector, whose element d_{jk} is the distance between the *j-th* neuron of the test face and the corresponding neuron of the *k-th* class.

Next, we convert the distance vectors into corresponding confidence vectors. One possible conversion method is to use a soft k-NN algorithm, which is briefly described as follows. The distance from the j-th neuron of the test face to its k-nearest neighbors are first arranged in increasing order: $d_{j1}^* \le d_{j2}^* \le \cdots \le d_{jk}^*$, then the confidence value for the k-th nearest neighbor is defined as,

$$c_{jk} = \frac{\log(d_{j1}^*+1)}{\log(d_{jk}^*+1)} \qquad (3)$$

It can be seen from E.q.3 that the class with minimum distance to the testing sub-block will yield a confidence value closer to one, while a large distance produces a very small confidence value, meaning that it is less likely for the testing sub-block to belong to that class.

Finally, the label of the test image can be obtained through a linearly weighted voting scheme, as follows,

$$Label = \arg\max_k \Big(\sum_{j=1}^{M} c_{jk}\Big), k = 1\ldots C \qquad (4)$$

IV. Experiments

A. Data Set and Methodology

To verify the performance of the proposed method, we have conducted various experiments on two well-known face image databases (AR and FERET). The AR database is employed to test the performance of the SOM-face algorithm under the conditions when partial occlusion and expression variant are involved. The FERET database is used to explore some

TABLE I.
DISTANCE MATRIX CALCULATION

Step1: Given C training images, the i-th image's BMU (Best Matching Unit) set is denoted by $Train_i = \{n_1^i \quad n_2^i \quad \cdots \quad n_M^i\}$, where $i = 1 \ldots C$, and M is the number of sub-blocks of each face.
Step2: Divide the test image x into M non-overlapping sub-blocks with equal size.
Step3: *1)For Single SOM-face strategy:* ♦ Project M sub-blocks of the test image onto the single trained map, the obtained BMU set is denoted by $Test_x = \{n_1^x \quad n_2^x \quad \cdots \quad n_M^x\}$. ♦ For $i=1 \ldots M$ (1)Calculate the distance between n_i^x in $Test_x$ and its counterparts n_i^k in $Train_k, k = 1 \ldots C$, (2)The resulting distances are denoted by $\{d_{i,k}, k=1 \ldots C\}$. *2)For MSOM-face strategy:* ♦ Project M sub-blocks of the test image onto the C trained maps separately, denote each of the obtained BMU set by $Test_x^k = \{n_1^x \quad n_2^x \quad \cdots \quad n_M^x\}$, $k = 1 \ldots C$. ♦ For $k = 1 \ldots C$: (1) In k-th map, calculate the distance between each n_i^x in $Test_x^k$ and its counterparts n_i^k in $Train_k$, $i=1 \ldots M$. (2) The resulting distances are denoted by $\{d_{i,k}, i=1 \ldots M\}$.
Step4: arrange the obtained distances into an $M \times C$ *distance matrix D*, with its element being $\{d_{i,k}, k=1 \ldots C, i=1 \ldots M\}$.

practical properties of the proposed algorithm, such as the selection of the size of sub-block.

Except when stated otherwise, all the experiments reported here were conducted as follows. In the localizing step, the training images were partitioned into non-overlapping sub-block with equal size. Then either a single SOM map or multiple SOM maps was trained in batch-mode using the obtained sub-blocks. The training process was divided into two phases as recommended by [25], that is, a rough training phase (to adjust the topological order of the weight vectors) and a fine-adjustment phase (to fine tune the feature map so as to provide an accurate statistical quantification of the input space). 100 updates were performed in the first phase, while 400 times in the second one. The initial weights of all neurons were set to the greatest eigenvectors of the training data, and the neighborhood widths of the neurons converged exponentially to 1 with the increase of training time. The soft kNN ensemble classifier described before was used for final classification decision.

Finally, to evaluation the experimental results of the SOM-face algorithm compared to other methods, we have conducted pairwise 1-tail statistical test using the method described in [47]. Recently, several other authors have also adopted the same statistical testing method [43, 49]

B. Experiments on the AR Database

The AR face database [44,45] contains over 4,000 color face images of 126 people's faces (70 men and 56 women), including frontal view faces with different facial expressions, illumination conditions, and occlusions (sun glasses and scarf). There are 26 different images per person, taken in two sessions (separated by two weeks), each session consisting of 13 images. In our experiments, a subset of 1200 images from 100 different subjects (50 males and 50 females) were used, which were the same dataset used by Martinez et al. in their experiments [18,21]. All the experimental results reported in this section concerning the local probabilistic approach were also quoted directly from [18] and [21]. Some sample images for one subject are shown in Fig.5.

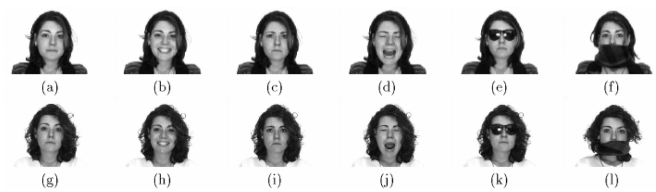


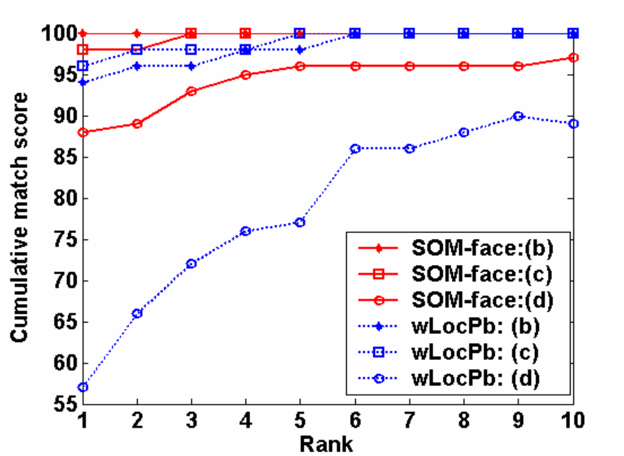
Fig.5. Sample images for one subject of the AR database

Before the recognition process, each image was cropped and resized to 120x165 pixels and then converted to gray-level images, which were then processed by a histogram equalization algorithm. Except when stated otherwise, in our experiments, the sub-block size of 5x3 was used and only the single SOM strategy was employed. According to [46], both the top 1 match and the top k matches were considered through the testing procedure.

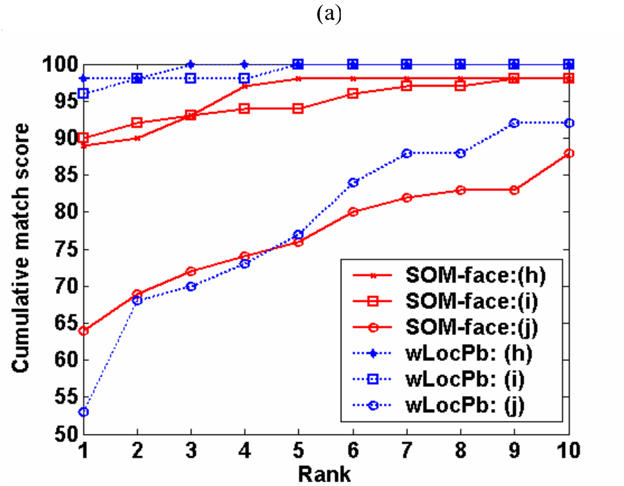
1) Variations in Facial Expressions

Firstly, we conducted experiments to investigate the classification capability of the proposed method under varying facial expressions. The results were compared with those of the weighted local probabilistic approach [18]. Specifically, in this series of experiments, the neutral expressions images (Fig.5a) of the 100 individuals were used for training, while the smile, anger and scream images in the first sessions (Figs.5b, 5c, and 5d) and those in the second sessions (Figs.5h, 5i, and 5j) were used for testing. Thus there were 300 testing images in all for each experiment.

Results of the two experiments are shown in Fig.6a and b, respectively, where the horizontal axis is rank and the vertical axis is cumulative match score, representing the percentage of correct matches with the correct answer in the top k matches. It can be seen from Fig.6 that, when the happy and angry images (Figs.5b, 5c, 5h, 5i) were presented to the system, the performance of the two methods is comparable. However, when the "screaming" faces (Figs.5d, 5j), which differ most from the training sample, were used for testing, the SOM-face algorithm achieved much higher performance than the weighted local probabilistic approach, to be more specific, 30% and 10% higher accuracy is yielded respectively concerning the *top 1 match rate*. Moreover, it should be noted that the weighted local probabilistic approach must use additional training samples to obtain the needed weighted information of



(a)



(b)

Fig.6. Comparative performance of the SOM-face algorithm and the weight local probabilistic approach (wLocPb) [18] with expression variant involved.

TABLE II

COMPARISON OF PCA, 2DPCA, AND SOM-FACE ALGORITHM CONCERNING THE TOP 1 RECOGNITION ACCURACY (%)

Index	PCA	2DPCA	SOM-face
(b)	97	99	100
(c)	91	94	98
(d)	63*	69*	88
(h)	67*	72*	88
(i)	63*	69*	90
(j)	40*	46*	64

The asterisks indicate a statistically significant difference between SOM-face and the compared method at a significance level of 0.05 in the trials.

each local area for each expression, while the SOM-face achieves comparable or better recognition performance by using only one reference facial image per individual. Therefore, the proposed method seems to have a higher degree of robustness to variant facial expressions.

For reference, we present the top 1 recognition rates one would obtain with the classical PCA and 2DPCA [43] in Table II.

2) Variations in partially occluded conditions

Next we want to study the applicability of the proposed method to partially occluded images, where both the simulated and the real occlusions were considered.

First the case of simulated occlusion was examined. The neutral expression images (Fig.5a) of the 100 selected individuals were used for learning, while the smiling, angry and screaming images with simulated partial occlusions were used for testing (Figs.5b, 5c, 5d). In each face, only those sub-blocks without being occluded were used for recognition. The occlusion was simulated by discarding some sub-blocks from the test face image, i.e. setting the gray values of all pixels within a sub-block to zeros. There were four modes of occlusions simulated in all by discarding the sub-blocks 10% each time, in the way of row by row from the bottom to top and vice versa (Figs.7b,7c), and column by column from left to right and vice versa (Figs.7d,7e), respectively. The results at every percentage of occlusion are denoted as occ_h in accordance with [18], with h={0,1,2,3,4,5}, for h=0, no portions of the image were occluded ; for h=1, only 10% of the images were occluded, etc.

(a) (b) (c) (d) (e)

Fig. 7: Illustration of the image in different occlusion modes. Image (a) is the testing image without occlusion, while images (b) through (e) are faces in four occlusion modes, respectively.

The results of the simulated-occlusion experiment are shown in Fig.8, where Figs.8 (a-d) display the corresponding results of the occlusion modes shown in Figs.7 (b-e), respectively. We can find that half face occlusion does not harm the performance of the recognition system much (see Fig.8a, 8c, 8d) except the occlusion of upper face (see Fig.8b). This observation agrees with the previous results obtained by Brunelli and Poggio [8], i.e., the use of the mouth area may lower the performance of the system. This can be explained by the fact that the lower half contains the mouth and cheeks, which can be easily affected by most facial expression variation. However, to what extent this effect affects all recognition techniques is still a problem needing to be further studied.

Next we examined the capability of the proposed method to handle real occlusions. For that purpose, two classical wearing occlusions, the sunglasses and the scarf occlusion (Figs.5e, 5f, 5k, 5l), were studied by conducting another experiment. As usual, the neutral expression images (Fig.5a) of the 100 individuals were used for training, while the occluded images (Figs.5e, 5f, 5k, 5l) were used for testing. Here both the training images and the testing images were not preprocessed by the histogram equalization algorithm to avoid the unwanted effect of the occluded portions of the faces, and the occluded portions of the image were not used for recognition. The results are shown in Fig.9.

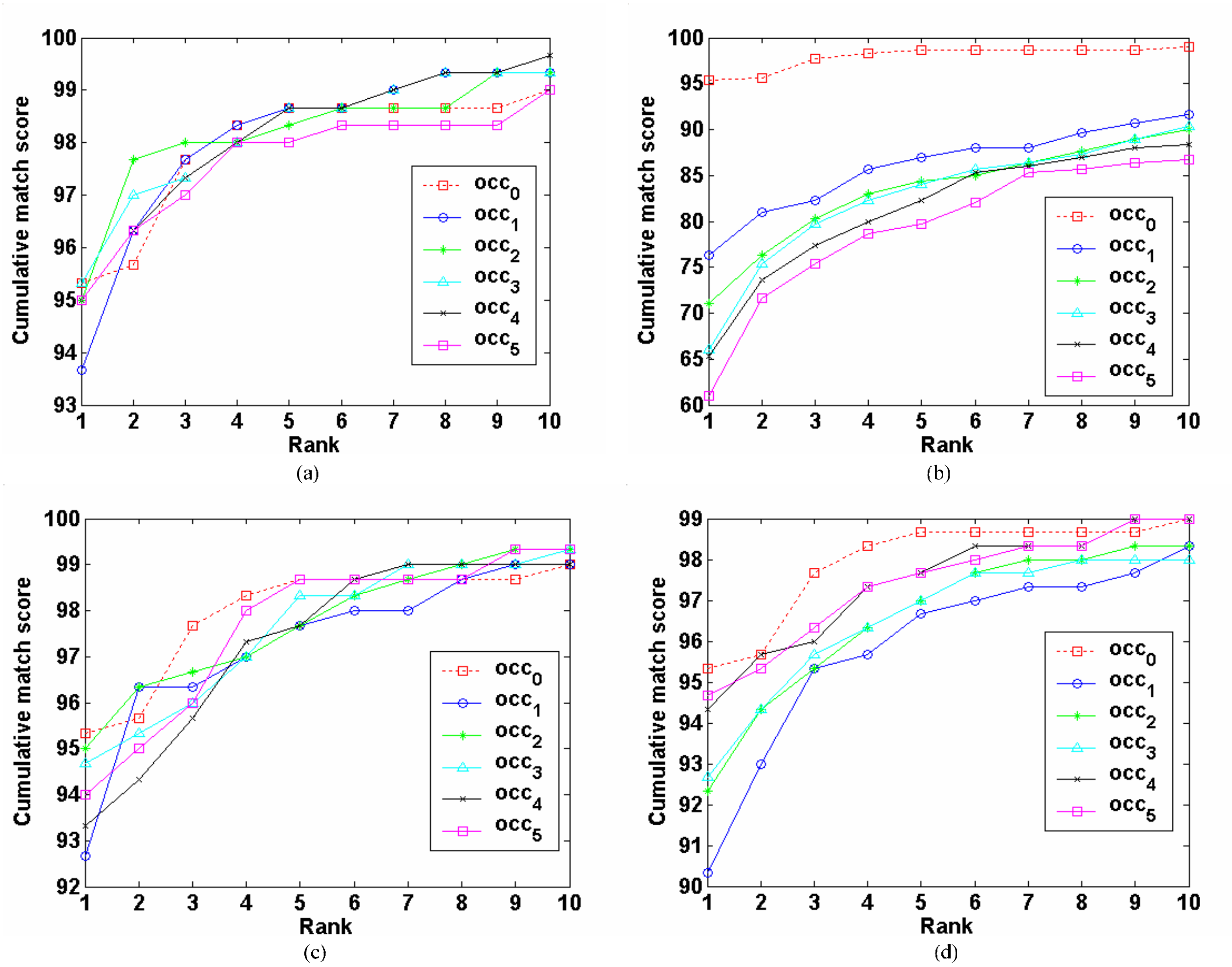


Fig.8. Performance of the SOM-face algorithm under different occlusion modes

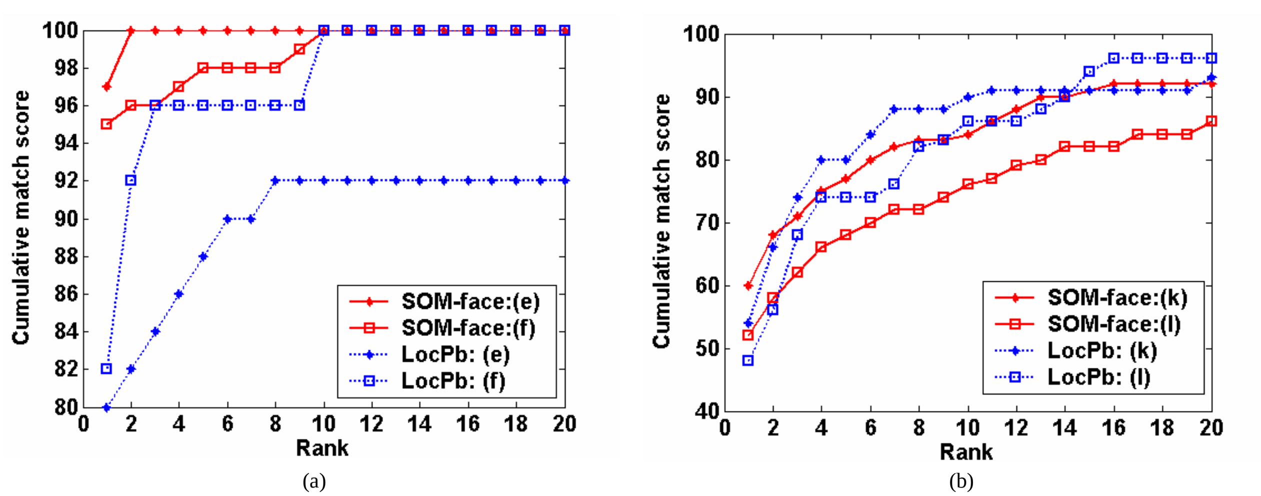
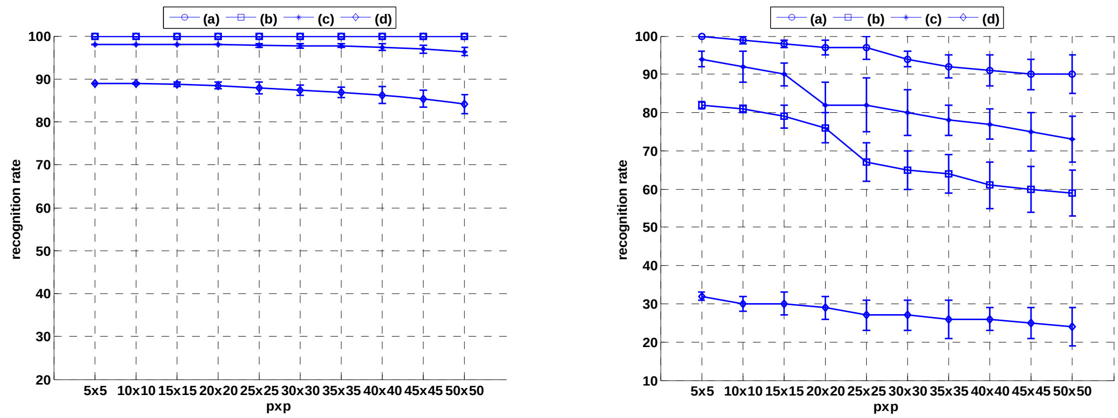


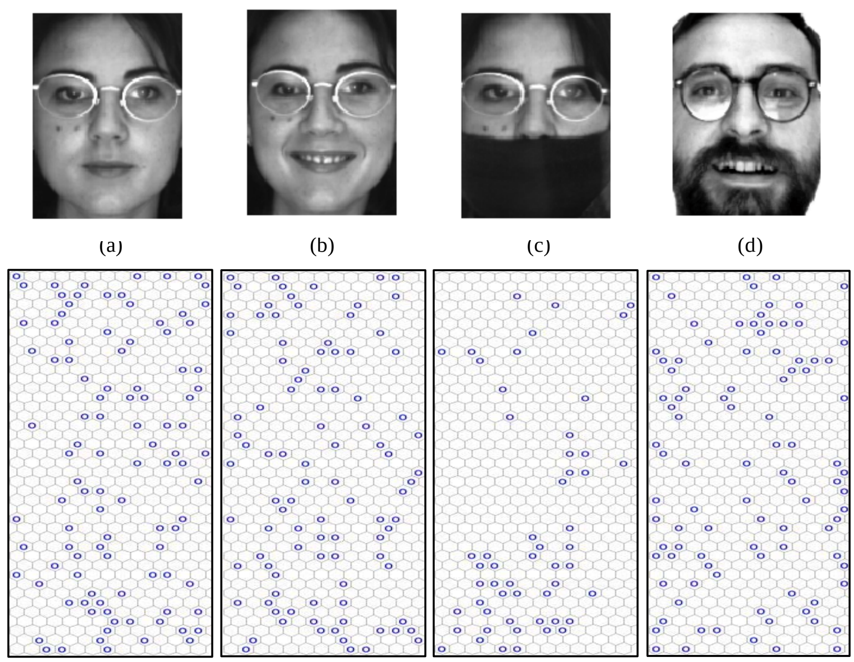
Fig.9. Comparative performance of the SOM-face algorithm and the local probabilistic approach (LocPb) with real occlusions involved

It can be seen from Fig.9 that SOM-face consistently achieved better recognition rate than the local probabilistic approach. Specifically, when the time factor is not involved (Figs.5e, 5f), the SOM-face outperformed the local probabilistic approach by about 10% or higher, concerning the *top 1 match rate*. It is interesting to note that when the duplicate images (Figs.5k, 5l) were presented to the system, the results of both algorithms reveal that the occlusion of the eyes area led to better recognition results than the occlusion of the mouth area as observed in [18]. This seems to be conflicting with our previous observation (Fig8b.) that upper part of the face is more important than the lower part. We think, however, that this contrary is a specific phenomenon, which is mainly caused by the fact that the scarf occluded each face irregularly and thus



(a) (b)

Fig.10. Recognition rate when the occlusion size and location is unknown. (a) the SOM-face algorithm and (b) the local probabilistic approach



(a) (b) (c) (d)

(e) (f) (g) (h)

Fig.11. An example of the corresponding distributions of the Best Matching Units (BMUs) of image (a-d) in the SOM topological space (e~h). (The size of the SOM map has been scaled for better display)

made the occluded face much more difficult to be identified than the face occluded by the sunglass (see Fig.5).

The occlusion experiments reported above assume that the system knows what is occluded and what is not beforehand. So how good our algorithm is when such an assumption is not valid? To answer this question, we have conducted another set of experiments as follows: the neutral expression images (Fig.5a) of the 100 individuals were used for learning, while the neutral, smiling, angry and screaming images for testing (Figs.5a-5d). To simulate the occlusion, we randomly localized a square of size pxp (5<p<50) pixels in each of the four testing image, with all the pixels inside the square set to zeros as done in [18]. Such synthetic occlusions were independently conducted 100 times, and the mean and standard deviation of the results for each of the facial expression groups were shown in Fig.10a, while the results one would obtain with the local probabilistic approach [21] were shown in Fig.10b for comparison. Fig.10 reveals that our method demonstrates highly robust performance against occlusions even when the occlusion size and location are unknown to the system. It can also be observed from Fig.10 that the results on Fig.5a and Fig.5b are both so perfect that they can not be visually distinguished from the figure. Moreover, our method consistently performs better than the local probabilistic approach, while the latter works much better than the classical PCA method [18].

3) *Visualization of the SOM-face*

Before ending the experiments on the AR database, we will exploit the visualization capability of SOM, which enables us gain some insight into the class distribution of high dimensional face image. In particular, the distribution of the BMUs of four images (three of which come from the same class, Fig.11a, 11b, 11c, one from some other class,Fig.11d) are visualized in the SOM map (Fig.11), and the map was learned using the neutral expression images (Fig.5a) of the 100 individuals, with the map size of 12x15 neurons for better display. Several observations can be made from Fig.11. Firstly, the figure shows that the projections of two matching images (Fig.11e, 11f) are not matched exactly in lower dimensional space. However, the similarity between the two can still be captured by the soft k-NN ensemble classifier for a correct recognition. Secondly, it can be observed that, in the feature space, the intra-class similarity between the two matching images (Fig.11e, 11f) is larger than the inter-class similarity between the two un-matching images (Fig.11e, 11h). How to enhance such a class-specific distribution will be an interesting topic of our future research. Thirdly, even when some of the sub-blocks are missing from the face (e.g., being occluded, Fig.11c), the left sub-blocks can still be used to identify the testing image (see Fig.11g, where only the best matching units of the sub-blocks not being occluded are displayed), that's why the SOM-face algorithm exhibits high robustness against missing information.

C. *Experiments on the FERET Database*

In this second series of experiments, a larger database than AR, i.e., FERET database is used [46]. The FERET database consists now of 13,539 facial images corresponding to 1,565 subjects, which are diverse across gender, ethnicity, and age. Several standardized subsets of FERET images have been defined, including a common set of gallery images (*fa*), and four different probe sets. In this study, the common gallery images *fa* are used for training, which consisted of images of 1196 people with one image per person, while *fb* probe set is selected for testing, which contains 1,195 images of subjects. only facial expressions variance is involved between the corresponding images in *fa* and *fb*. Some samples of the database are shown in Fig.12. Before the recognition process, the raw images were normalized and cropped to a size of 60×60 pixels.

Fig.12. Some raw images in the FERET database

On this dataset, five experiments were conducted to further inspect some practical aspects involved in the SOM-face algorithm, such as choosing an appropriate value for the *k* value of the soft *k*-NN ensemble classifier, defining the size of sub-blocks, and so on. We note that in the implement of the MSOM-face strategy, alternative version of way to calculate the distance matrix is used due to its better recognition performance, i.e., combine all small maps to one big map and then calculate the distance matrix as the single SOM-face

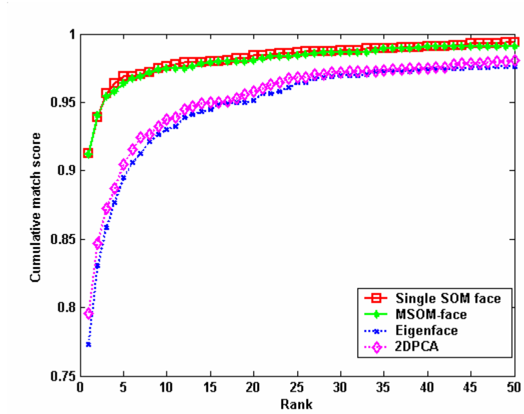


Fig.13. Comparative performance of the single SOM-face strategy, MSOM-face strategy, Eigenface and 2DPCA algorithm on FERET

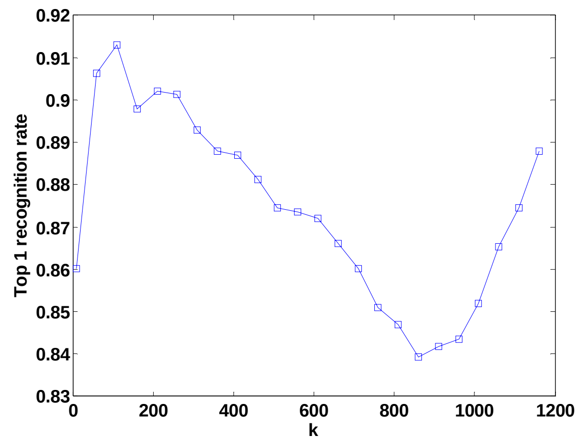


Fig.14. Top 1 recognition rate of the single SOM-face algorithm with varying k-values

strategy.

Experiment 1 Firstly, the performance of the two SOM-face based algorithms (single SOM-face and MSOM-face) on the subset was evaluated. All the images were divided into sub-blocks of 3x3. The results are shown in Fig.13.

It can be observed from Fig.13 that the results of the single SOM-face strategy and the MSOM-face strategy were close, whereas both outperformed the standard Eigenface technique and 2DPCA by 10%-15%, respectively. Pairwise 1-tailed statistical test indicates that such difference is statistically significant at 0.95 level of significance.

Experiment 2 Secondly, we investigated the problem of choosing an appreciate *k*-value for the soft *k*-NN classifier. For this purpose, a sub-block size of 3x3 was used and then a single SOM map was trained using all 1196 images in the gallery set. And the top *1* recognition rate of the system on 1195 probe images is measured, using the soft *k*-NN ensemble classifier with varying *k*-values. The results are shown in Fig.14. The figure reveals that it would be appropriate to set the *k*-value in the range of 10% to 20% of the size of the database, while the larger *k*-value did not necessarily produce better results. Nevertheless, as an exception to the above remark, the largest *k*-value that could be set, i.e., the size of the training samples, might be an interesting point worthy to be examined.

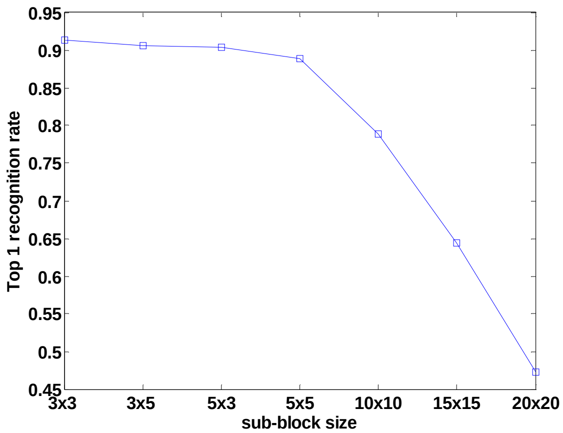


Fig.15. Top 1 recognition rate of the single SOM-face algorithm with varying sub-block size

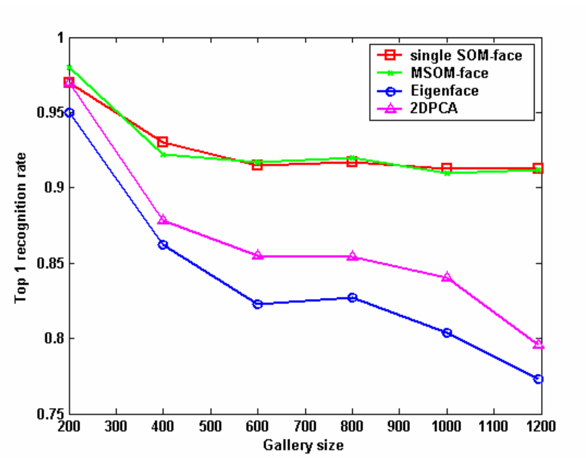


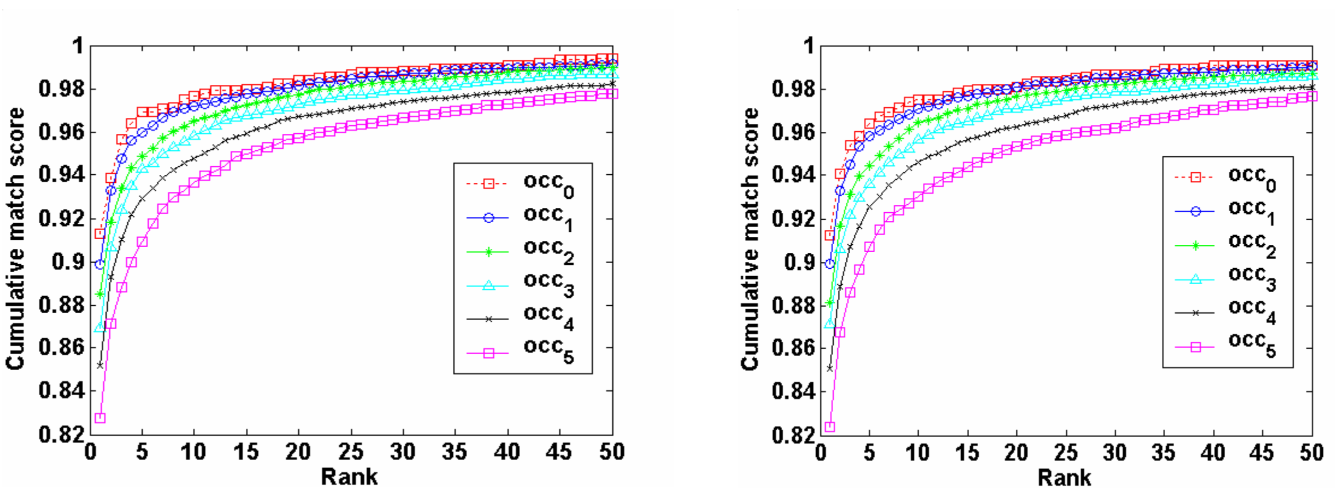
Fig.16. Top 1 recognition rate of the single SOM-face, MSOM-face, Eigenface and 2DPCA as a function of gallery size

Experiment 3 The effect of different sub-block sizes on the performance of the proposed algorithm was also studied. We partitioned the training images and testing images to various sub-block sizes (e.g., 3x3, 3x5, etc.) and, then, repeated the above experiment (i.e., Experiment 2, but with fixed *k*-value) under different sub-block sizes. The results are displayed in Fig.15. It can be seen that smaller sub-block seems be more beneficial to the performance of the system. Intuitively, the choice of sub-block size reflects the balance between generalization and specialization, that is, as the sub-blocks gets smaller, the degree of generalization grows higher, while the degree of specialization becomes lower.

Experiment 4 The computational complexity is considered then. Formally, let the number of training faces and testing faces be n_{tr} and n_{test}, respectively. Suppose that each face is divided into *M* sub-blocks with *d* dimensions each sub-block. Also let the number of iterations be *T*, the number of neurons in the self-organizing map be n_c, and the number of neurons used in the neighborhood calculation be n_h, respectively. Then the computational complexity of the single SOM-face strategy consists of two terms, i.e., $O(n_{tr}Mdn_hn_cT)+O(n_{test}Mkd)$, where the first term stems from the training of SOM map [39], while the second term results from the soft *k* nearest neighbors ensemble classifier.

The computational complexity of the multiple SOM-face strategy can be calculated in the same way as the single SOM-face strategy, except that a third term will be added, i.e. $O(n_{tr}Mn_cd)$, which results from the recalculating of the prototype vectors when a new face is added into the database. However, the training of the MSOM-face is much faster than that of the single SOM-face due to its smaller training set and smaller SOM map. On our IBM xSerises 235 server with two Intel® Xeon™ processors and 1GB memory, the time to train a single map of 3472 neurons for 1196 faces, with 3x3 sub-block and 500 iterations, took about 36.1 hours. However, only 1.6 hours was needed if the MSOM-face strategy was used to train 1196 small maps with 100 neurons each.

To investigate the incremental learning capability of the MSOM strategy, another experiment was conducted using different gallery sizes. Fig.16 shows the top 1 recognition rates of different methods as the function of gallery size.



(a) (b)

Fig.17.summaries of the performance of the SOM-face algorithm at different percentage of occlusions, (a) single SOM-face algorithm (b) MSOM-face algorithm.

Experiment 5 Finally, we repeated one of the simulated occlusion experiments done on the AR dataset (see Fig.7). In this experiment, all the images in gallery set were used for learning, while all the images in *fb* probe set were artificially partially occluded according to the same way as described in Section4.2.2 (Fig.7.). The sub-blocks size was taken as 3x3 and both single SOM-face and MSOM-face strategies were tried. Only the averaged results at different percentages of occlusions are displayed in Fig.17. Once again, the results illustrate the robustness of the proposed method against partial occlusions.

V. CONCLUSIONS

In this paper, we introduce a very simple but effective method called "SOM-face" to address the problem of face recognition with one training image per person. The images were allowed to vary in expressions and have partial occlusion. The proposed method has several advantages over some of the previous methods such as weighted local probabilistic method [18], the standard Eigenfaces technique [27] and the 2DPCA algorithm [43]. Firstly, it can achieve comparable or better performance than the mentioned methods with no single extra virtual samples needed to be generated. Secondly, it shows higher robustness against expression variance and partial occlusions. Thirdly, this method is very intuitive due to the visualization capability of SOM, which enables us gain some insight into the class distribution of high dimensional face image.

We attribute these advantages to the seamless connection between the three parts of the method: By localization of the training samples, the robust performance against global changes is increased. By the unsupervised and nonparametric learning of the SOM, the similarity relationship of the sub-blocks in the input space is preserved in the SOM topological space. And finally, by the soft k-NN ensemble classifier, the similarity relationship is captured and exploited to enhance the whole system's robustness.

However, it is worth mentioning that the proposed method assumes that the system knows what is occluded and what is not occluded in advance. It's naturally to ask the question how to decide the location and size of the occlusion. In a general case, this is a very difficult problem since any facial portions can be occluded in any shape at any grey level. One possible way to deal with this problem is to train several specific feature detectors corresponding to each facial part (e.g. eyes, nose, mouth and profile) [7] so as to detect the non-occluded facial parts and use them for recognition purpose. Wu et al [6] have recently presented a method to automatically locate the region of eyeglasses if the system is given a face wearing eyeglasses, using an offline trained eye region detector. Other methods such as active appearance models (AAM [5]), deformable shape models [41] are also useful. Another way to deal with the problem is to try to bypass the problem with the help of users, for example, Gutta et al discarded the whole half of the occluded face and only used the information from either left or right half of the test face for recognition [42]. Since the first way is not so mature and heavily dependent on the accuracy of the specific feature detectors, we prefer to use the second method currently, i.e., incorporating man in loop. Although experiments show that our method demonstrates highly robust performance against occlusion even when the occlusion size and location is unknown to the system, the possible need of the manual effort of the user can be regarded as a drawback of the proposed system. Thus, further studies along the first line will be the focus of our future work.

Finally, the proposed method could also be regarded as a general paradigm for dealing with *small sample problem*, in which the training set is transformed and enlarged by being partitioned into multiple sub-blocks. This paper shows that this paradigm works well in the scenario of face recognition with one training image per person. It is anticipated that this method is also effective in scenarios where each person has two (or more, but still *small sample*) training images.

VI. ACKNOWLEDGMENTS

We want to express our gratitude towards the anonymous reviewers whose valuable comments and suggestions greatly improved this paper. We also want to thank Dr. Aleix M. Martinez for providing us the AR database. Portions of the research in this paper use the FERET database of facial images collected under the FERET program.

REFERENCES

[1] R. Chellappa, C. L. Wilson and S. Sirohey, "Human and machine recognition of faces: A survey," *Proc. of the IEEE*, 83(5):705~740, 1995
[2] J. Daugman, "Face and gesture recognition: Overview," *IEEE Trans. Pattern Analysis and Machine Intelligence*, 19(7):675~676, 1997.
[3] W. Zhao, R. Chellappa, P. J. Phillips, and A. Rosenfeld, "Face Recognition: A Literature Survey, "*ACM Computing Survey*, December Issue, pp. 399-458, 2003
[4] A. K. Jain, Y. Zhong, and S. Lakshmanan, "Object matching using deformable templates," *IEEE Trans. Pattern Analysis and Machine Intelligence* 22(1):4-37,2000.
[5] T. F. Cootes, G. J. Edwards, and C. J. Taylor, "Active appearance models", ECCV, 2 (1998), 484–498.
[6] C.W, C.Liu, et al, "Automatic Eyeglasses Removal from Face Image". *IEEE Trans. Pattern Analysis and Machine Intelligence* , vol. 26, no. 3, pp. 322-336, 2004.
[7] H.A. Rowley,S.Baluja,and T.Kanade, "Neural Network-Based Face Detection", *IEEE Trans. Pattern Analysis and Machine Intelligence* , vol. 20, no. 1,Jan.1998.
[8] R. Brunelli, and T. Poggio, "Face recognition: Features versus templates." *IEEE Trans. Pattern Analysis and Machine Intelligence*, 15(10),1042-1062. 1993.
[9] A. J. O'Toole and H. Abdi, "Low-dimensional representation of faces in higher dimensions of the face space." *Journal of Optical Society of America*, 10(3): 405-411, 1993.
[10] A. K. Jain,R.P.W.Duin,and J. Mao, "Statistical Pattern Recognition:A Review,"IEEE Trans. *Pattern Analysis and Machine Intelligence*, 22(1),4-37, 2000.
[11] S.Lin; S.Kung; L. Lin," Face recognition/detection by probabilistic decision-based neural network". *IEEE Trans. on Neural Networks* ,8(1),114-132, 1997.
[12] J. Lu, K. N. Plataniotis, and A. N. Venetsanopoulos, "Face Recognition using kernel direct discriminant analysis algorithms." *IEEE Trans. on Neural Networks* ,14(1),117-126, 2003.

[13] D. L. Swets, J. Weng, ”Using discriminant eigenfeatures for image retrieval.” *IEEE Trans. on Pattern Analysis and Machine Intelligence* 18(8),831-836,1996。
[14] P. N. Belhumeur, J. P. Hespanha, and D. J. Kriegman, ” Eigenfaces vs. Fisherfaces: recognition using class specific linear projection,” *IEEE Trans. on Pattern Analysis and Machine Intelligence* 19(7), 711-720, 1997
[15] I. Lapidot; H.Guterman; A.Cohen, “Unsupervised speaker recognition based on competition between self-organizing maps”, *IEEE Trans. on Neural Networks*,13(4),877-887, 2002
[16] B. Moghaddam, T. Jebara., A. Pentland, “Probabilistic visual learning for object detection,” *Proc Int Conf Computer Vision*, pp. 786–793, 1995
[17] A. M. Martinez and A. C. Kak, ”PCA versus LDA,” *IEEE Trans. on Pattern Analysis and Machine Intelligence*,vol.23.no.2,pp.228-233,Feb.2001.
[18] A. M. Martinez, “Recognizing imprecisely localized, partially occluded, and expression variant faces from a single sample per class,” *IEEE Trans. Pattern Analysis and Machine Intelligence* 25(6): 748-763,2002.
[19] S. Chen, D. Zhang, and Z.-H. Zhou. ”Enhanced $(PC)^2A$ for face recognition with one training image per person,” *Pattern Recognition Letters*, Vol25,1173-1181,2004.
[20] D. Beymer and T. Poggio, "Face Recognition from One Example View," Science 272(5250), 1996.
[21] A. M. Martínez, “Recognition of Partially Occluded and/or Imprecisely Localized Faces Using a Probabilistic Approach,” Proc. of IEEE Computer Vision and Pattern Recognition (CVPR), Vol. I, pp. 712-717, 2000.
[22] A. M. Martinez, ”Matching Expression Variant Faces,” *Vision Research*,Vol.43,Issue 9,pp1047-1060,2003.
[23] S.-H. Lin, S. Y. Kung, and L.-J. Lin, “Face recognition/detection by probabilistic decision-based neural network,” *IEEE Trans. Neural Networks*,vol. 8, pp. 114–132, Jan. 1997
[24] T. Kohonen, “Self-Organization and Associative Memory,” 2nd edition, Berlin: Springer-Verlag,1988
[25] T. Kohonen, “Self-Organizing Map,“2nd edition, Berlin: Springer-Verlag,1997.
[26] B. Zhang, H. Zhang, and S. Ge, “Face Recognition by Applying Wavelet Subband Representation and Kernel Associative Memory”, *IEEE Trans. on Neural Networks*, 15(1), 166-177, 2004
[27] L Sirovich and M. Kirby, “Low Dimensional Procedure for the Characterization of Human Faces,” *JOSA-A(4)*, No. 3, pp. 519-524, 1987.
[28] A. Pentland, B. Moghaddam, and T. Starner, “View-based and modular eigenspaces for face recognition.” *Proceedings of the IEEE International Conference on Computer Vision and Pattern Recognition*, Seattle, WA, 84-91, 1994
[29] D. Valentin, H. Abdi, et al, “Connectionist models of face processing: a survey,” *Pattern Recognition*,v27,no.9,pp.1209-1230,1994.
[30] P. R. Villeala, J. uan Humberto Sossa Azuela, “Improving Pattern Recognition Using Several Feature Vectors,” *Lecture Notes in Computer Science*, Springer-Verlag Heidelberg,2002.
[31] L. I. Kuncheva, C. J. Whitaker, ”Feature subsets for classifier combination: an enumerative experiment,” In: Kittler, J., Roli, F. (eds.): *LNCS*, Vol. 2096. Springer, Berlin 228-237, 2001.
[32] Y. Linde, A. Buzo, R. M. Gray, ”An algorithm for vector quantizer design, “*IEEE Trans. Commun.*, Vol. COM-28,pp.84-95,Jan.1980.
[33] W. Pedrycz, “Fuzzy sets in pattern recognition: Methodology and methods, “*Pattern Recognition*, vol.23,no.1/2,pp.121-146,1990.
[34] S. Lawrence, C. L. Giles, A. Tsoi, and A. Back, “Face recognition: A convolutional neural-network approach.” *IEEE Trans. on Neural Networks*, vol.8, no.1, 98-113, 1997
[35] C. Liu and H. Wechsler: "Evolutionary Pursuit and Its Application to Face Recognition", *IEEE Trans. Pattern Analysis and Machine Intelligence* , vol. 22, no. 6, pp. 570-582, 2000.
[36] L. Wiskott, J. M. Fellous, N. Kruger and C. V. D. Malsburg, “Face Recognition by Elastic Bunch Graph Matching,” *Intelligent Biometric Techniques in Fingerprint and Face Recognition*, L.C.Jain Et al. eds.,Springer-Verlag,1999.
[37] M. Fleming and G. Cottrell, “Categorization of faces using unsupervised feature extraction,” In Proc. IEEE IJCNN International Joint Conference on Neural Networks, 1990.
[38] J. D. McAuliffe, L. E. Atlas, and C. Rivera. “A comparison of the LBG algorithm and Kohonen neural network paradigm for image vector quantization,” *Proc. ICASSP-90,Int.Conf. on Acoustics,Speech and Signal Processing*, Vol.IV,pp.2293-2296,1990.
[39] T .Kohonen, S. Kaski, K. Lagus, et al, “Self Organization of a Massive Document Collection, “*IEEE Transactions on Neural Networks*, Vol.11 No.3, May 2000.
[40] J. A. Kangas, T. Kohonen, J. T. Laaksonen, “Variants of self-organizing maps,” *IEEE Transactions on Neural Networks*, 1:93－99, 1990.
[41] T.K.Leung,M.C.Burl,and P.Perona, ”Finding Faces in Cluttered Scenes Using Random Labeled Graph Matching”, Proc. Fifth IEEE int’l Conf. Computer Vision,pp.637-644,1995.
[42] S.Gutta, H.Wechsler,” Face Recognition Using Asymmetric Faces”, David Zhang, Anil K. Jain(Eds.);ICBA2004,LNCS 3072,pp162-168,2004.
[43] J. Yang, D. Zhang, A. F. Frangi & J. Yang, "Two-dimensional PCA: A new approach to appearance-based face representation and recognition," *IEEE Trans. Pattern Analysis and Machine Intelligence*, vol. 26, no.1, 131-137,2004.
[44] A. M. Martínez and R. Benavente, “The AR Face Database,” *CVC Technical Report*,no.24,June 1998.
[45] A. M. Martínez and R. Benavente, “The AR Face Database,” http://rvl1.ecn.purdue.edu/~aleix/aleix_face_DB.html,2003.
[46] P. J. Phillips, H. Wechsler, J. Huang, P. J. Rauss, “The FERET database and evaluation procedure for facerecognition algorithms.” *Image and Vision Computing*, 16(5), 295-306, 1998
[47] W. Yambor, B. Draper, and R. Beveridge, “Analyzing PCA-Based Face Recognition Algorithms: Ei-Genvector Selection and Distance Measures,” *Empirical Evaluation Methods in Computer Vision*, H. Christensen and J. Phillips, eds., Singapore: World Scientific Press, 2002.
[48] P. S. Penev and J. J. Atick “Local Feature Analysis: A General Statistical Theory for Object Representation,” *Network: Computation in Neural Systems*, vol.7,no.3, pp.477-500,1996.
[49] J.Kittler,M.hatef,R.Duin,and J.Matas. “On Combining classifiers”. *IEEE Trans. Pattern Analysis and Machine Intelligence* , vol. 20, no. 3, pp. 226-239, 1998.
[50] R.Fergus,P.Perona, and A.Zisserman. “Object class recognition by unsupervised scale-invariant learning”. In *Int,Conf. Computer Vision and Pattern Recognition*, Vol 2, pp 264-275, Madision, Wisconsin,USA,June 2003.

Xiaoyang Tan received the B.Sc. and M.Sc. degrees in computer science from Nanjing University of Aeronautics & Astronautics, China, in 1993 and 1996, respectively. Currently he is a Ph.D .student at the Department of Computer Science & Technology, Nanjing University. His research interests are in machine learning, pattern recognition and computer vision.

Songcan Chen received the B.Sc. degree in mathematics from Hangzhou University (now merged into Zhejiang University) in 1983. In Dec. 1985, he completed the M.Sc. degree in computer applications at Shanghai Jiaotong University and then worked at Nanjing university of Aeronautics & Astronautics in Jan. 1986 as an assistant lecturer. There he received a Ph.D. degree in communication and information systems in 1997. Since 1998, as a full professor, he has been with the Department of Computer Science and Engineering at NUAA. His research interests include pattern recognition, machine learning and neural computing. In these fields, he has authored or coauthored over 70 scientific journal papers.

Zhi-Hua Zhou (S'00-M'01) received the B.Sc., M.Sc. and Ph.D. degrees in computer science from Nanjing University, China, in 1996, 1998 and 2000, respectively, all with the highest honor. He joined the Department of Computer Science & Technology of Nanjing University as a lecturer in 2001, and is a professor and head of the LAMDA group at present. His research interests are in artificial intelligence, machine learning, data mining, pattern recognition, information retrieval, neural computing, and evolutionary computing. In these areas he has published over 40 technical papers in refereed international journals or conference proceedings. He has won the Microsoft Fellowship Award (1999), the National Excellent Doctoral Dissertation Award of China (2003), and the Award of National Outstanding Youth Foundation of China (2004). He is an associate editor of *Knowledge and Information Systems* (Springer), and on the editorial boards of *Artificial Intelligence in Medicine* (Elsevier) and *International Journal of Data Warehousing and Mining* (Idea Group). He served as the organizing chair of the 7th Chinese Workshop on Machine Learning (2000), program co-chair of the 9th Chinese Conference on Machine Learning (2004), and program committee member for numerous international conferences. He is a senior member of China Computer Federation (CCF) and the vice chair of CCF Artificial Intelligence & Pattern Recognition Society, a councilor of Chinese Association of Artificial Intelligence (CAAI), the vice chair and chief secretary of CAAI Machine Learning Society, and a member of IEEE and IEEE Computer Society.

Fuyan Zhang is a professor at the Department of Computer Science & Technology of Nanjing University, China. His research interests are in information processing and multimedia. In these areas he has published over 100 technical papers in refereed international journals or conference proceedings. He is currently the chair of the Multimedia Computing Institute of Nanjing University.