Face Recognition under Occlusions and Variant Expressions with Partial Similarity

Xiaoyang Tan, *Member* Songcan Chen, Zhi-Hua Zhou, *Senior Member*, Jun Liu

***Abstract*— Recognition in uncontrolled situations is one of the most important bottlenecks for practical face recognition systems. In particular, few researchers have addressed the challenge to recognize non-cooperative or even uncooperative subjects who try to cheat the recognition system by deliberately changing their facial appearance through such tricks as variant expressions or disguise (e.g. by partial occlusions). This paper addresses these problems within the framework of similarity matching. A novel perception-inspired non-metric partial similarity measure is introduced, which is potentially useful in deal with the concerned problems because it can help capturing the prominent partial similarities that are dominant in human perception. Two methods, based on the general *golden section* rule and the *maximum margin* criterion, respectively, are proposed to automatically set the similarity threshold. The effectiveness of the proposed method in handling large expressions, partial occlusions and other distortions is demonstrated on several well-known face databases.**

***Index Terms*— Similarity measure, pattern recognition, machine learning, partial similarity, non-metric similarity, face recognition, self-organizing map (SOM)**

I. Introduction

DUe to its wide applications in information security, law enforcement and surveillance, smart cards, access control, and others, face recognition technique has received significantly increased attention from both the academic and industrial communities during the past several decades [43]. However, face recognition in uncontrolled situations remains one of the most important bottlenecks for practical face recognition systems [23].

The goal of this paper is to deal with one class of face recognition problem where some of facial appearances in a given face image are badly deformed by such variations as large expression changes or partial occlusions (or disguise) due to sunglasses, scarves,mustaches and so on. Such variations in facial appearance are commonly encountered in uncontrolled situations and may cause big trouble to the face-recognition-based security system but are less studied in literatures [5]. Notice that in this paper, we don't intend to deal with other commonly encountered variations in uncontrolled conditions like lighting changes and ageing effect, which are of interest but usually change people's facial appearance in a more holistical way. By contrast, the facial appearance changes caused by variant expressions and partial occlusions are mostly local in nature, i.e., only parts of facial appearance change largely while others are less affected. The challenge lies in that such local deformations or occlusions in facial appearance can be anywhere and in any size or shape in a give face image and we don't have any prior knowledge about it.

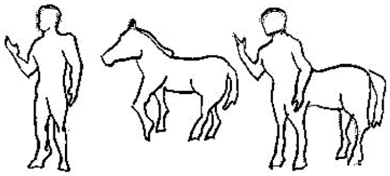


Fig. 1. An illustration of the intuition of partial similarity measures [9]

This paper proposes to address this from the aspect of partial similarity matching by exploiting the spatial contiguousness nature of occlusions and other local deformations. This idea can be best illustrated with Fig. 1, where many observers will feel that both the person and the horse are similar to the centaur, but the person and the horse are not similar to each other at all. Why? One possible reason is that when comparing two images, human beings tend to "focus on the portions that are very similar and are willing to pay less attention to regions of great dissimilarity" [9]. Inspired by this, one goal of this paper is to design an effective mechanism to support such a robust perception of similarity by humans in face recognition systems, i.e., automatically detecting and capturing the significant partial similarities between two face images while ignoring the unreliable and unimportant features due to expression changes, occlusions or disguises.

Besides a naive *golden section* rule [29]-based method, we achieve this within the framework of maximum margin theory by focusing on the learning of optimal partial similarity thresholds. The optimized criterion proposed for threshold learning is discriminative in nature, and is more like a combination of a set of regularized local likelihood functions rather than a single global objective function as in the usual large margin settings. This means that we don't aim to seek a single hyperplane that separate the data into different class as well as possible but to fit a set of local discriminative models corresponding to each class. Since it is unlikely to obtain a closed form solution in our setting, an effective greedy search strategy is developed to optimize the values of parameters. The resulting algorithm is able to automatically identify the most discriminative local parts for recognition without making any assumption on the distributions of the deformed facial regions (for example, we don't need to know where the face images are occluded before recognition), and has the advantage of being intuitive and simple to implement.

The rest of this paper is organized as follows. Section II reviews some related works. Section III proposes our partial similarity method. The problem of how to set the similarity

Corresponding author: Tel: +86-25-8489-2805; fax: +86-25-8489-3777. E-mail: s.chen@nuaa.edu.cn (S. Chen); zhouzh@nju.edu.cn (Z.-H. Zhou); x.tan@nuaa.edu.cn (X.Tan).

threshold is addressed in Section IV. To validate the feasibility and effectiveness of the proposed method, extensive experiments are conducted and reported in Section V. Finally, Section VI concludes and raises several future issues.

II. Related Work

This section briefly reviews related face recognition algorithms.

Most classical face recognition methods seek an optimal representation subspace in which the unwanted influence of facial variations is reduced as much as possible. Typical methods in this line include most subspace methods and their kernelized versions (e.g., PCA, [28], [35], LDA, [3], Kernel PCA plus LDA [15] [41]), Bayesian intra/extrapersonal classifier ([19]) and other manifold-based methods such as Laplacianface (LPP, [6]). Technically, these methods gain their goals by relying on some linear or nonlinear transformations on the holistic image vectors used for training, and are shown to be robust against 'global' variations such as lighting or ageing effect [15]. However, they may not fit well with those images with large local deformations such as occlusions or disguise, partly due to that the resulting holistic representations are usually far deviated from the normal patterns.

There are also some holistic methods that designed to handle this by imposing various local constraints on the transformations of basis. For example, in Independence Component Analysis (ICA, [2]) architecture I, it is required that the learned holistic basis should be statistically independent to decorrelate the second order statistics, which is similar to the Local Feature Analysis (LFA, [22]) technique but with different startpoint. The Local Non-negative Matrix Factorization (LNMF, [13]) attempts to seek a set of local basis which are additive and sparse when used for representation. Inspired by the idea of compressed sensing, a more recent method (SRC, [37]) also tries to exploit the sparsity nature of the occluded face images by minimizing the L1 norm of coefficients.

Alternative methods assume that the pattern of the partial occlusions or local deformations are contiguous, hence a convenient way to use this characteristic is to partition the image into blocks and model each type of block separately. In the Weighted Local Probabilistic Subspace method (WLPS, [16]), a mixture Gaussian model is used for that purpose, which is extended later in [31] with an unsupervised neural network to represent the subspace of local features. Such methods works well in some situations but when the available samples are relatively few, the discriminativity of each block is of importance. Recently, Singh et al. [25] proposed to use discriminative 2D log polar Gabor phase features to distinguish people with disguise using only one single image per person [32].

Ivanov et al. [8] introduced a 'semi-local' method in which various components such as eyes, mouth and nose are first detected by separate SVM classifiers, and then a new (partial) face image is "reconstructed" with these components, which is further fed into another SVM classifier for final recognition. This method is similar to ours but it implicitly assumes that all of the interested components can be reliably identified and especially should not be occluded, such that a good 'reconstruction' could be obtained. Instead, we adopt a strategy that essentially allows any part of the given facial image to be deformed or occluded, by focusing on the learning of optimal discriminative thresholds, which are helpful in deciding which part's weight should be decayed during recognition, thus relaxing the assumption in Ivanov et al's method.

III. Partial Distance Measure for Face Recognition

The overall architecture of the proposed method is shown in Fig.2. To enable the capture of the partial similarity and the integration of the spatial information, face images are partitioned into local facial regions (sub-blocks) at first. Then, all the sub-blocks are mapped into an SOM topological space to obtain a compact and robust representation [31], where the nearest neighbor search is performed using the proposed partial distance measure, and the training face image with the smallest partial distance to the probe face image is selected to give the final identity.

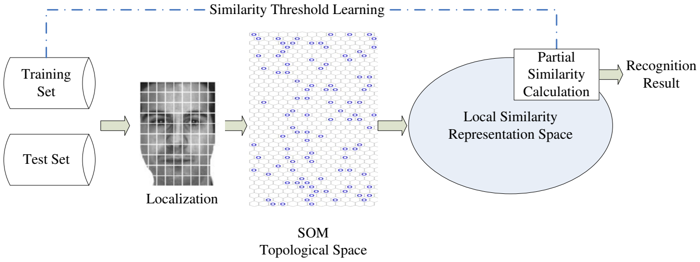

Similarity Threshold Learning
Training Set
Test Set
Localization
SOM
Topological Space
Partial Similarity Calculation
Local Similarity Representation Space
Recognition Result

Fig. 2. Overall architecture of the proposed method

A. Face Image Localization

Among all the potential definitions of local facial region, perhaps the simplest is the one that defines a local facial region as a rectangle or sub-block in the image. In particular, each face image is divided into $K(= dim_a/dim_b)$ non-overlapping sub-blocks with equal size, where dim_a and dim_b are the dimensionality of the whole image and the sub-block, respectively. Typical sub-block size used in practice is 4×4 pixels. Although complex local feature descriptors such as Gabor wavelet [15], Local Binary Pattern (LBP) [34], Local Gabor Binary Patterns Histogram Sequence (LGBPHS) [42], etc. exist, the simple gray-level-based descriptor is used in this paper. That is, each sub-block is represented as a local feature vector (LFV) by concatenating the pixels in the sub-block. Such a gray-level-based local feature representation has been shown useful in face detection [36] and face recognition [12], [31] before. Actually, the obtained LFVs could not only encode the appearances of the local regions but also preserve the spatial configurations of 2D face images to some extent.

B. The Local Pairwise Similarity Matrix

After defining the local regions of face images, the local pairwise similarity matrix D between a probe face $\boldsymbol{q}$ and every face $\boldsymbol{x}_i$ in the training set T can be constructed. Denote the set of sub-blocks of $\boldsymbol{q}$ and $\boldsymbol{x}_i$ as $\{\boldsymbol{q}^k\}_{k=1}^K$ and $\{\boldsymbol{x}_i^k\}_{k=1}^K$ respectively, where K is the number of sub-blocks. Then, each element of D can be calculated as $d_{ki} = d(\boldsymbol{q}^k, \boldsymbol{x}_i^k)$, which is the local pairwise distance between the two corresponding (i.e., the *k*-th) sub-blocks from the probe face and the *i*-th training face. This is a convenient way to exploit the inherent regularity of face image by implicitly taking the spatial layout information into account.

Note that although face images usually span a non-metric manifold from the holistic view, it is commonly assumed that the neighborhood relationship of local patches of face images can be well approximated with metric distance [26]. Hence,

$$d_{ki} = d(\boldsymbol{q}^k, \boldsymbol{x}_i^k) = \|\boldsymbol{q}^k - \boldsymbol{x}_i^k\|_p \tag{1}$$

where $\|\cdot\|_p$ is the L_p norm defined on the LFVs, i.e. $\boldsymbol{q}^k$ and $\boldsymbol{x}_i^k$, with $p \geq 1$. Then, the similarity matrix can be written as:

$$D(\boldsymbol{q}, T) = \begin{bmatrix} d(\boldsymbol{q}^1, \boldsymbol{x}_1^1) & \cdots & d(\boldsymbol{q}^1, \boldsymbol{x}_N^1) \\ d(\boldsymbol{q}^2, \boldsymbol{x}_1^2) & \cdots & d(\boldsymbol{q}^2, \boldsymbol{x}_N^2) \\ \vdots & \vdots & \vdots \\ d(\boldsymbol{q}^K, \boldsymbol{x}_1^K) & \cdots & d(\boldsymbol{q}^K, \boldsymbol{x}_N^K) \end{bmatrix} \tag{2}$$

where N is the number of training faces in *T*. The matrix shown in Eq. 2 plays an important role in the following computation. More specifically, each column of D encodes the overall distribution of the local similarity information between the probe face and a training face. In terms of the similarity approaches [21], this matrix is actually a local similarity representation of the data set, containing all the information needed for the subsequent recognition task.

The global distance $d(\boldsymbol{q}, \boldsymbol{x}_i)$ between the probe face $\boldsymbol{q}$ and the training image $\boldsymbol{x}_i$ is computed as follows:

$$d(\boldsymbol{q}, \boldsymbol{x}_i) = \sum_{k=1}^{K} d_{ki} = \sum_{k=1}^{K} d(\boldsymbol{q}^k, \boldsymbol{x}_i^k) \tag{3}$$

That is, the sum of the local pairwise distances. If and only if the L_1 norm is used to calculate the local pairwise distances, Eq. 3 degenerates to the traditional L_1 norm between $\boldsymbol{q}$ and $\boldsymbol{x}_i$. While for other local distance measures such as the L_2 norm, the situation is different. Generally, if the L_p norm with $p > 1$ were used, to make the definition of the global distance $d(\boldsymbol{q}, \boldsymbol{x}_i)$ consistent with the traditional L_p norm, Eq. 3 can be changed to:

$$d^p(\boldsymbol{q}, \boldsymbol{x}_i) = \sum_{k=1}^{K} (d_{ki})^p \tag{4}$$

Here only the L_1 norm is considered for simplicity but without loss of generality.

C. The Partial Distance Measure

If an appropriate threshold τ is given, the set of local pairwise distances $\{d_{ki}\}_{k=1}^K$ can be divided into two subsets, that is,

$$S = \{k | d_{ki} \leq \tau, k = 1, \ldots, K\} \tag{5}$$

$$F = \{k | d_{ki} > \tau, k = 1, \ldots, K\} \tag{6}$$

The *S* and *F* are called the *similar* subset and *dissimilar* subset, respectively. The threshold allows one to adjust the proportion of the similar/dissimlar sub-blocks of two face images. For example, a large threshold will lead to that more sub-blocks are assigned to the *similar* subset.

Subsequently, Eq. 3 can be re-written as:

$$d(\boldsymbol{q}, x_i) = \sum_{k} d_{ki} = \sum_{k \in S} d_{ki} + \sum_{k \in F} d_{ki} \tag{7}$$

That is, the global distance of two face images is equal to the sum of the local pairwise distances of similar portions and dissimilar portions. Eq. 7 can be generalized to:

$$d_{PD}(\boldsymbol{q}, \boldsymbol{x}_i, \beta) = \beta \sum_{k \in S} d_{ki} + (1 - \beta) \sum_{k \in F} d_{ki} \tag{8}$$

where $\beta \in [0, 1]$ is a parameter which balances the contribution of the similar and dissimilar portions. The value of this parameter can be defined based on the statistical distribution information of the similar and dissimilar portions. Here β is defined to be:

$$\beta = min(1, \frac{|S|}{|F|}) = min(1, \frac{|S|}{K - |S|}) \tag{9}$$

where $|S|$ and $|F|$ are the number of similar sub-blocks and dissimilar sub-blocks, respectively. Clearly, β is in the interval [0 1], and its value has positive relationship with the number of similar sub-blocks due to the relatively higher importance of similar sub-blocks compared to dissimilar sub-blocks.

In the rest of this paper, the generalized distance defined in Eq. 8 is called *partial distance* (PD) to emphasize the contribution of partial facial regions. A crucial problem remains to be tackled in the calculation of the PD distance is how to properly set the threshold τ, which will be addressed in Section IV.

Two variants of the PD distance are explored here. The first is calculated by summing the first l most similar pairwise sub-block distances while the second simply counts the number of similar sub-blocks between two images.

1) Continuous Partial Distance (cPD): Sort the pairwise sub-block similarity matrix of Eq. 2 in increasing order column by column, and consider only the first l local distances of each column. Formally, denote the elements of the *i*-th column of the similarity matrix as $\{d_{ki}\}_{k=1}^K$, and sort them in increasing order such that:

$$d_{ki} \leq d_{li}, \forall\ 1 \leq k < l \leq K. \tag{10}$$

Now the similarity threshold can be defined as *l*, and the continuous partial distance measure is:

$$d_{cPD}(\boldsymbol{q}, \boldsymbol{x}_i, \beta) = \beta \sum_{k \leq l} d_{ki} + (1 - \beta) \sum_{k > l} d_{ki} \tag{11}$$

where β is set according to:

$$\beta = min(1, \frac{l}{K - l}), l \neq K \tag{12}$$

Actually, the PD distance shown in Eq. 8 is also a kind of continuous partial distance. The major difference between the PD distance and the *c*PD distance lies in the way of setting the similarity threshold, that is, *c*PD uses l instead of τ.

2) Discrete Partial Distance (dPD): The pairwise sub-block distance is first discreteized as follows:

$$I(d_{ki}) = \begin{cases} -1, & d_{ki} \le \tau \\ 0, & d_{ki} > \tau \end{cases} \tag{13}$$

Then, Eq. 8 becomes:

$$\begin{aligned} d_{dPD}(\boldsymbol{q}, \boldsymbol{x}_i, \beta) &= \beta \sum_{d_{ki} \le \tau} I(d_{ki}) + (1-\beta) \sum_{d_{ki} > \tau} I(d_{ki}) \\ &= \beta \sum_{d_{ki} \le \tau} I(d_{ki}) \end{aligned} \tag{14}$$

where β is as same as that in Eq. 9 and the last step is due to the definition of Eq. 13. Eq. 14 indicates that the global distance between two faces completely depends on the weighted number of similar sub-blocks.

D. Using the Partial Distance Measure

The identity of the probe face image can be obtained by the nearest neighbor rule using the partial distance:

$$label(\boldsymbol{q}) = \arg\min_{i=1\ldots N}(d_{PD}(\boldsymbol{q}, \boldsymbol{x}_i, \beta)) \tag{15}$$

where $label(\boldsymbol{q})$ is the class label of the probe face image $\boldsymbol{q}$, and $d_{PD}(\boldsymbol{q}, \boldsymbol{x}_i, \beta)$ can be calculated using either Eq. 8 or its variants such as Eq. 11 or Eq. 14.

E. Properties of the Partial Distance Measure

The partial distance defined above has several appealing properties for face recognition tasks.

First, it automatically selects the most similar portions between two faces for comparison, which makes the complex intra-personal distribution being more compact.

Second, by definition, a distance measure is a metric distance if it satisfies four metric axioms, i.e., non-negativity, self-similarity, symmetry, and the triangle inequality [27]. If any of these axioms is violated, the concerned distance measure is called non-metric distance. It is obvious that the following two properties are satisfied by the partial distance:

1) non-negativity: $d_{PD}(\boldsymbol{q}, \boldsymbol{x}_i) \ge 0$
2) symmetry: $d_{PD}(\boldsymbol{q}, \boldsymbol{x}_i) = d_{PD}(\boldsymbol{x}_i, \boldsymbol{q})$

Nevertheless, the partial distance measure defined above is not transmissive, i.e., it violates the triangle inequality (transitivity should be followed from the triangle inequality [9]). This occurs mainly because different sub-blocks can make contributions in different comparisons. As for the centaur example shown in Fig. 1, the similar sub-blocks between person and centaur and those between horse and centaur are different.

Moreover, it is worth mentioning that in the discrete version of the partial distance measure (dPD), two sub-blocks can be regarded as similar even if they are not identical, given that the pairwise distance between them is below some predefined threshold (Eq. 13). Such a property somewhat violates the self-similarity axiom (i.e., $d(q, x_i) = 0$ if and only if $q = x_i$). However, it should be noted that this property actually increases the tolerance against slight local distortions and potentially increases the possibility of finding the correct matching for a given face.

IV. Similarity Threshold Setting

This section will deal with the problem of how to properly set the similarity threshold τ (see Eqs. 5 and 6). Two strategies are proposed, i.e. the *golden section* strategy and the maximum marginal-based strategy.

A. Setting Similarity Threshold Based on the Golden Section Rule

The *golden proportion* or *golden section* is a harmonic way of dividing a segment with length into two parts. Two quantities are said to be in the golden ratio if "the whole (that is, the sum of the two parts) is to the larger part as the larger part is to the smaller part" [38]. This number is a powerful empirical value widely used in natural science, not only because of its simplicity but also because "the golden section has redundance and stability which allow self-organized systems to be organized" [29].

Recall that the face image $\boldsymbol{x}$ has been partitioned into two portions, i.e. the similar portion S (Eq. 5) and the dissimilar portion F (Eq. 6). If such a partition is golden proportion-compliant, then:

$$\frac{total\ size\ of\ \boldsymbol{x}}{total\ size\ of\ S} = \frac{total\ size\ of\ S}{total\ size\ \ ofF} \quad \text{hence}$$

$$\begin{aligned} total\ size\ of\ S &= \frac{\sqrt{5}-1}{2} \times (total\ size\ of\ \boldsymbol{x}) \\ &= 0.618 \times (total\ size\ of\ \boldsymbol{x}). \end{aligned} \tag{16}$$

That is, similar portion takes about 61.8% of the whole face image. Obviously, this can be used to guide the setting of the similarity threshold for the cPD distance. Due to its empiricism in nature, further discussion will be postponed to the experimental section (Section V-B).

B. Learning Similarity Threshold Based on the Maximum Margin Criterion

Intuitively, good thresholds should be class-dependent in nature, i.e., different thresholds should be set for different person (class). In order to ensure a small generalization error, the threshold should also assign a face image to the correct class with high confidence, which is usually called *margin* in literature. However, in the practice of large margin, such as in [1] and [39], the margin over the whole training set is usually needed to be optimized, hence imposing a relatively strong constraint on the cost function. In this paper, an alternative strategy which does not require all the similar samples to be clustered tightly at the same time, is adopted. That is, the margin of each class is optimized separately. This local strategy not only makes the optimization task become easier, but also allows to obtain a series of optimal class-dependent thresholds, one for each class. The overall effect is that the samples from each class are closely clustered, respectively.

Formally, let y_i denote the class label of the training example $\boldsymbol{x}_i$. The index set of the training examples belonging to the c-th class is $H_c = \{i | y_i = c, i = 1, \ldots, N\}$, and the index set of the examples from other class is $\overline{H}_c = \{i | y_i \neq c, i = 1, \ldots, N\}$. Then the training set of the c-th class is denoted as $X_c = \{\boldsymbol{x}_i, i \in H_c\}$ with size $|H_c|$. Furthermore, denote the threshold of the c-th class as τ_c.

The optimization process for τ_c involves a Leave-One-Out validation strategy on X_c. First, fetch one example $\boldsymbol{x}_i \in X_c$ as the validation example, and all the remaining examples in the training set T as prototypes. Then, try to label this validation example using the partial distance under some given threshold τ_c (with Eq. 15). Suppose that the classification result be $LOO(\boldsymbol{x}_i, \tau_c)$, the average margin of the c-th class is then defined as:

$$\begin{aligned}\overline{m}_c(X_c, \tau_c) = \frac{1}{|H_c|} \sum_{i \in H_c} \{ & 1(LOO(\mathbf{x}_i, \tau_c) = y_i) \\ & [\min_{j \in \overline{H}_c} d_{PD}(\mathbf{x}_i, \mathbf{x}_j, \tau_c) - \min_{\substack{j \in H_c \\ j \neq i}} d_{PD}(\mathbf{x}_i, \mathbf{x}_j, \tau_c)]\}\end{aligned} \tag{17}$$

where $1(u)$ is the indicator function which takes 1 if u is true and 0 otherwise. Eq. 17 says that if a training example $\boldsymbol{x}_i$ $(i \in H_c)$ is correctly classified in the Leave-One-Out validation, then the classification confidence can be measured by the margin between the nearest training example of other classes (the first term) and the nearest prototype (except $\boldsymbol{x}_i$ itself) of the c-th class (the second term). Clearly, only positive values of $\overline{m}_c(X_c, \tau_c)$ express correct classifications, and the larger the value, the higher the classification confidence.

Eq. 17 is then used as the cost function to be optimized in the training phase, and its output will be the needed class-dependent similarity threshold τ_c^*. However, maximizing a margin function like Eq. 17 is generally difficult. Here a straightforward greedy search strategy is adopted. Considering that the distance value range in the local similarity matrix D (Eq. 2) may vary from a probe image to another, we first scale each element of D in the range of [0, 1] such that the greedy search can be done in a manageable range. A local linear transformation named continuous histogram equalization normalization is employed in this paper. The major advantage of this local transformation is that each distance value tends to be less influenced by the large value in the whole value range of D, compared to traditional (global) transformation. On the other hand, if the values of distance matrix are exponentially distributed, one can also use logarithmic normalization for the same purpose.

Finally, after the similarity thresholds for all the classes are learned, they can be used for face recognition. A majority voting scheme [10] is employed for that purpose in this paper:

1) For a probe face $\boldsymbol{q}$, calculate the similarity matrix $D(\boldsymbol{q}, T)$ and normalize every element of D as described above.
2) Estimate the partial similarity between the probe face and all the training examples under thresholds of each class (with Eq. 8 or Eq. 14), respectively. At each time, label the probe face as the identification of the training example that has the maximum partial similarity to the probe (Eq. 15).
3) Use a standard majority voting strategy to find the winner class, which gives the final identification of the probe face.

Except when stated otherwise, the similarity threshold of the original PD distance and its variant dPD is set according to this strategy.

C. Embedding with the Self-Organization Maps

One problem remains to be tackled is that the direct calculation of the pairwise distances in the input space may become computational intensive when the number of sub-blocks at hand is large. One way to deal with this problem is to map, or embed the local facial vectors into a low-dimensional embedding space such that [7]: (1) the distances of the embedded vectors approximate the actual distances, and (2) the similarity matrix computation can be performed in the less intensive embedding space.

In this paper, the Self-Organizing Maps (SOM, [11]), as one of the most efficient and effective techniques that satisfy the above two requirements at the same time, is adopted. Notice that other options, such as locally linear embedding(LLE, [26]), locality-preserving projection (LPP, [6]), and multidimensional scaling (MDS, [33]), can also be naturally used. The SOM is a two-layered network with its output layer commonly being a two-dimensional grid (lattice) (Fig.3). Each neuron (node) of the lattice also stores a weight vector (also called *codebook* or *reference vectors*), which in turn defines a Voronoi region in the input space. A Voronoi region associated with a weight vector is a set of points closer to that vector than any other (Fig.3). By training, the SOM algorithm can produce a topological ordering of the feature map in the input space in the sense that neurons that are adjacent in the lattice will tend to have similar weight vectors [11]. This is one of the most important properties of the SOM, hence the SOM grid is commonly called the topological space. More than that, each Voronoi region actually defines a deformable subspace in the input space, since all the sub-blocks falling in the same Voronoi region would be finally mapping to the same neuron in the SOM grid (i.e., the Best Matching Unit (BMU), Fig.3). This feature is valuable to improve the system's tolerance against slight local distortions (e.g., random pixel corruption or misalignment) in the underlying Voronoi region.

The SOM training process can be done offline. After that, 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. Then, one can perform the local similarity matrix computation based on that information. As an example (Fig.3), the distance $d(b1, b2)$ between two sub-blocks $b1$ and $b2$ is approximated by the distance between their corresponding BUMs ($a1$ and $a2$, respectively) in the 2D SOM grid, i.e., $d(a1, a2)$, where $d(\cdot, \cdot)$ is a pre-defined distance measure.

V. Experimental Results

In this section, a series of experiments are carried out to evaluate the effectiveness of the proposed partial similarity methods.[1]

[1] A MATLAB implementation of the proposed algorithm is available at http://cs.nju.edu.cn/zhouzh/zhouzh.files/publication/annex/PD.htm.

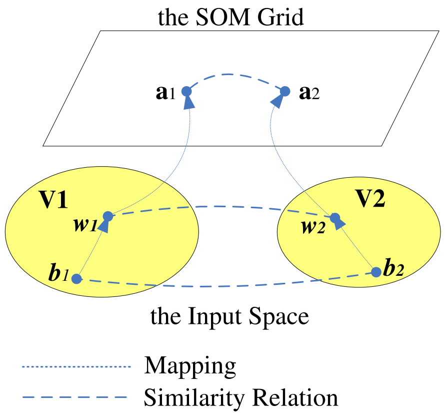


Fig. 3. Illustration of SOM embedding. $a1$ and $a2$ are two nodes in the SOM grid, whose weight vectors are $w1$ and $w2$, respectively, and their corresponding Voronoi regions are V1 and V2, separately. $b1$ and $b2$ are two sub-block vectors in the input space, which are first mapping to $w1$ and $w2$ then to $a1$ and $a2$, respectively, such that the similarity relationship between $b1$ and $b2$ in the input space can be approximated in the two-dimensional SOM grid.

A. Databases and Experimental Settings

Three well-known face databases (AR [17], FERET [24] and ORL [3]) are used in this work. The AR database is the only real database available that contains images with large expression changes and disguise accessories, hence particularly suitable for this study, while the ORL is a classic database that contains such commonly-encountered distortions as slight variations in pose angle, glasses and alignment. The FERET database is much larger than the other two, and is used to test the performance of our method with both expression and occlusion variations. Next we give some description about the AR and ORL databases, and the description to FERET can be found in section V-E.

The AR face database [17] contains over 4,000 color face images of 126 people's faces (70 males and 56 females), including frontal view faces with different facial expressions, illumination conditions, and occlusions (with 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 1,200 images from 100 different subjects (50 males and 50 females) with frontal illumination are used, which is the same dataset used by Martinez et al. [16], [18].Some sample images for one subject are shown in Fig. 4.

The ORL database [3] contains 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 grayscale with a resolution of 92 × 112 pixels. Fig. 5 shows five raw images of two persons in this database.

In all the experiments except those on the ORL database, the original images are first normalized (in scale and orientation) such that the two eyes are aligned at the same position (the coordinates of the centers of the eyes are provided by the owner of the specific databases as metadata). Then, the facial areas are cropped from the face image. Finally, the cropped face areas are processed by a histogram equalization algorithm to reduce the influence of possible illumination variations. The sizes of each cropped image in the AR ,FERET and ORL database are 80×60, 80×60 and 92×112 pixels, respectively, with 256 gray levels per pixel. Notice that no registration or preprocessing is made on the images of the ORL database such that the algorithm's robustness against imprecise alignment can be tested.

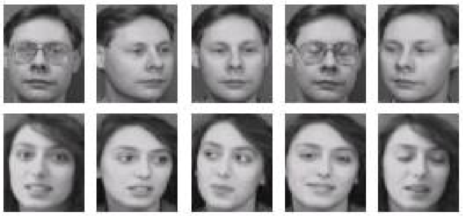


Fig. 5. Examples of images of two subjects in the ORL database.

To evaluate the performance of the compared methods, we have conducted pairwise one-tail statistical test under significance level 0.05, which has been popularly used in previous research [4], [40], [41].

B. A Pilot Experiment on the Importance of the Similar Portions

Since the basic assumption of the proposed method is that the similar portions are of high significance in comparing two images, it is meaningful to validate this assumption first. For this purpose, a pilot experiment is designed by increasingly including more similar portions for recognition. Meanwhile, the same number of randomly selected sub-blocks are also used for testing. It can be expected that if the assumption is valid, the proposed similarity-based method should work significantly better than the method using randomly selected sub-blocks.

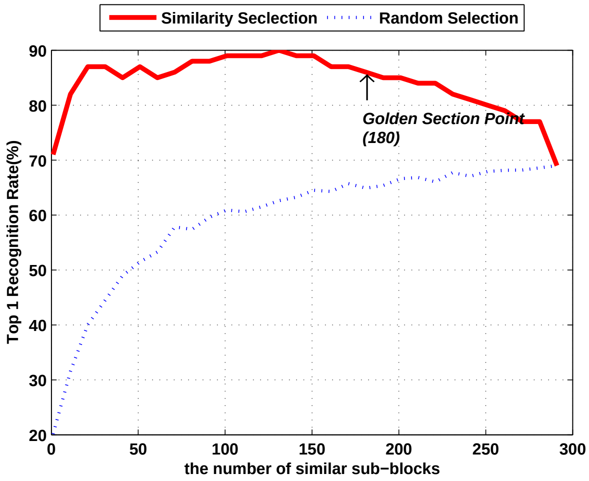


Fig. 6. The performance as a function of the number of similar portions used for recognition.

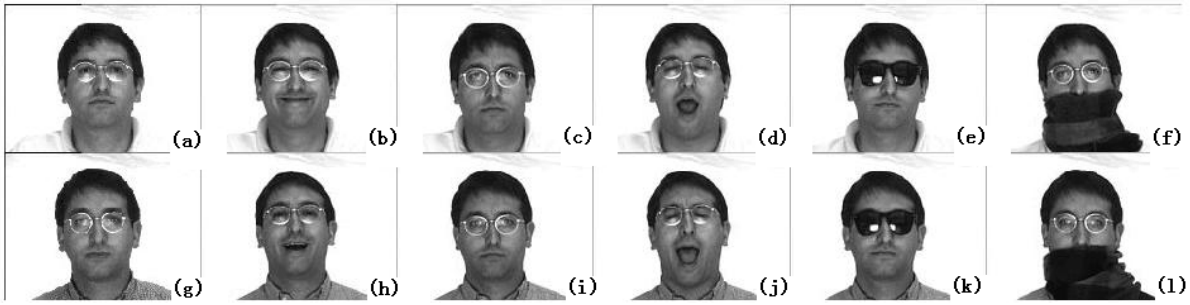


Fig. 4. Images of one subject in the AR database with four expressions and two partial occlusions. (a-f) are images taken from the first session while (g-l) the second session.

In the experiment, the neutral expression (Fig.4(a)) in the AR database is used for enrollment while the scream expression (Fig.4(d)) for testing. The reason for choosing scream expression for testing is that this expression causes the largest variation on the face appearance compared to other expressions. We evaluate the performance with the nearest neighbor strategy and the test for random selection is repeated for 50 times. The top 1 average matching rate is depicted in Fig. 6 as a function of the number of sub-blocks used for recognition, and the following observations can be obtained from this figure:

1) The difference between the performance of random and similarity selection is obvious. In particular, at the golden proportion point, the difference is over 20.0% and is statistically significant ($Z = 3.5$). This reveals that the similar portions play a crucial role in robust visual face image matching.
2) With the similarity-based selection strategy, the recognition system is able to rapidly reach its performance peak with the a few similar sub-blocks. After that, adding more sub-blocks would not be much helpful. Moreover, as more and more sub-blocks with larger pairwise distance are used, the performance curve begins to drop, which suggests that the performance curve suffers from the increasing noise and transformation errors.
3) Nevertheless, the tendency of the performance drop can be prevented if the threshold is properly set. It appears that the golden selection is a good choice for that purpose. To make this clear, the golden section point on the curve is marked in Fig. 6. It can be found that a good performance of 87.0% can be achieved at the golden proportion point. Experiments on other databases also exhibit the similar tendency.

In summary, this experiment suggests that similar portions are of high importance for robust face image comparison, and the golden proportion appears to be a good similarity threshold in practice although it may not be the best choice. Indeed, as described in Section IV-B, we have designed a more "discriminative" method to capture the optimal similarity for each class (each class consists of the faces obtained from a single individual) based on the maximum margin criterion. In the next sections, we will focus on investigating the feasibility and effectiveness of the latter strategy, and more results concerning the former strategy (i.e., the golden proportion-based one) will also be presented.

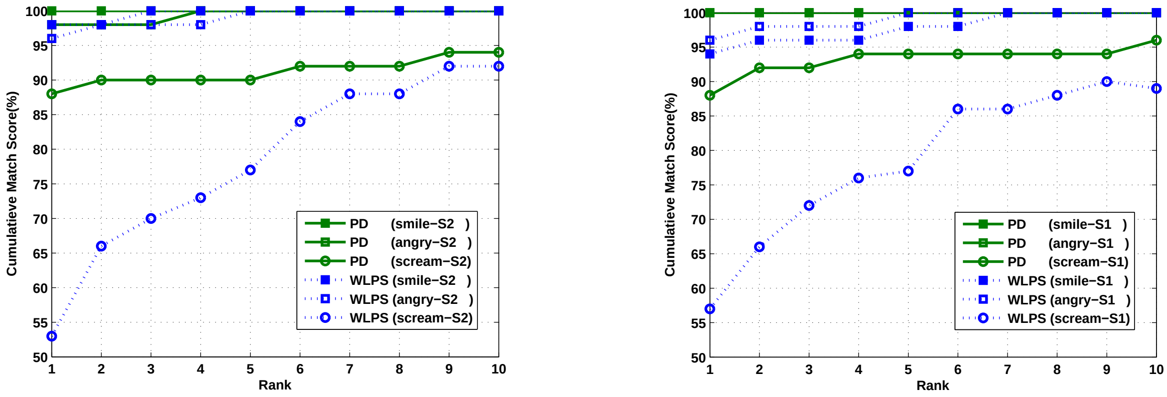
C. Variations in Facial Expressions

Three experiments are conducted in this section: the first two correspond to the cases where the group of people does not change (i.e., images used for training and testing belong to the same individual but without overlapping) and where the group of people does change (i.e., the subjects in the training set and testing set are different.), respectively. The third experiment tests the behavior of the recognition system when a specific "template" expression is left out from the gallery set. For the first two experiments in this section, we randomly partition the 100 persons into two groups with 50 persons each.

In the first experiment, for each person from the first group, the four expression-variant images taken in the first session (Fig.4(a)-(d)) are used to learn the similarity thresholds, while the testing is performed completely on the images taken in the second session(Fig.4(g)-(j)). More specifically, for each person in the same group, his/her image with neutral expression in the second session (Fig.4(g)) is used as enrollment data and the other images in the second session (Fig.4(h)-(j)) corresponding respectively to the smile, anger, and scream expressions, are used as testing images.

The results are shown in Fig. 7(a), where the horizontal axis is the rank and the vertical axis is cumulative match score, representing the percentage of correct matches with the correct answer in the top k matches. We have not included the results of cPD and dPD in this figure since their results are too similar to those of the PD method to be visually distinguishable. It can be seen that the proposed method is almost insensitive to the smile and angry expressions, and is quite robust against the extreme expressions such as the scream in the AR database. In addition, our method consistently performs better than the WLPS [16] method, and the difference in performance between the proposed method and WLPS is statistically significant over the scream expressions ($Z = 5$). This experiment also suggests that the similarity learned from some images of an individual can be generalized to other images from the same person.

Next, we want to investigate that whether the learned similarity can be generalized across the persons in the same database, where large cultural differences between groups do not exist. For that purpose, a completely new group of



(a)

(b)

Fig. 7. Results on expression variations on the AR Databases for (a) the first experiment and (b) the second experiment. S1: images taken in the first session; S2: images taken in the second session.

individuals is employed for testing, where the image with neutral expression of each person taken in the first session (Fig. 4(a)) is used as enrollment data and the other images in the same session corresponding respectively to the smile, anger, and scream expressions (Fig. 4(b),(c) and (d)), are used as testing images. Note that, we do not re-learn the optimal similarity thresholds for this new group of people, but just employing those obtained in the previous experiment. The results are shown in Fig. 7(b).

Fig. 7(b) reveals that using the similarity threshold learned from other individuals can still yield good performance of 100.0% for both simile and angry expression and near 90.0% for the scream expression. This finding suggests that the proposed method can effectively extract useful prior information from other groups for robust recognition, given that large cultural differences do not exist between groups.

Under the same setting of the second experiment, we compare our methods with several related methods, including NNC(Nearest Neighborhood Classifier with Euclidean distance), GKLDA(Kernel LDA with Gabor feature, [15]),ICA [2], LNMF [13], SRC [37], LFA [22], WLPS [16]. The results are shown in Table I. It can be observed that our methods perform better than most of the compared methods. The improvement is significant in cases involving difficult expressional variations like screaming.

Finally, to further verify the effectiveness of the proposed method in coping with unseen expressions, a leave-one-expression-out testing procedure [18] is employed. For example, if a happy expression is to be tested, the neutral, angry and scream expression will be used for training. This procedure is repeated two times in the AR database. In the first time, only those images taken during the first session are used; In the second time, the images of the first session are used for training, while those of the second session for testing. That means when the happy faces of the second session is used for testing, the smile, angry and scream faces of the first session are used for training.

The results are summarized in Table II. It can be observed that our methods exhibit better performance than the compared methods when a specific expression is "missing" from the templates. This mainly owes to the ability of our methods in capturing the prominent intra-personal partial similarity in a discriminative manner. While many other local methods such as LNMF [13], and ICA [2] are unsupervised in nature, they do not select the most useful local features for recognition. WLPS effectively employs the discriminative information from the training set with the mixture Gaussian model to improve its performance. In contrast, our method is both supervised and non-parametric, without making any assumption about the shape of the class distribution.

TABLE I
THE DETAILED RESULTS ON THE AR DATABASE(%)

	Expressions Variations			Partial Occlusions		
Algorithm	Smile	Angry	Scream	Glasses	Scarf	Average
NNC [20]	96.3	88.9 *	57.0 *	48.1 *	3.0 *	58.7 *
GKLDA	100.0	98.0	88.0	68.0 *	60.0 *	77.4 *
SRC	100.0	98.0	82.0	95.0 *	46.0 *	66.0 *
LNMF [20]	94.8	76.3 *	44.4 *	18.5 *	9.6 *	48.7 *
ICA	96.0	95.0	51.0 *	69.0 *	37.0 *	69.6 *
LFA [5]	96.0	92.0	76.0 *	10.0 *	81.0 *	71.0 *
WLPS [16]	96.0	96.0	56.0 *	80.0 *	82.0	82.0 *
cPD	100.0	100.0	86.0	98.0	84.0	93.6
dPD	100.0	100.0	86.0	97.0	92.0	95.0
PD	**100.0**	**100.0**	**88.0**	**98.0**	**90.0**	**95.2**

* The asterisks indicate a statistically significant difference between the compared method and the PD method at a significance level of 0.05.

D. *Variations in Partially Occluded Conditions*

In this section, we investigate the robustness of our methods with respect to partial occlusions. The AR database contains two classical wearing occlusions, i.e., the sunglasses and the scarf occlusions (see Fig. 4(e)(f)).

In particular, the neutral expression images from 100 persons (Fig. 4(a)) are used as enrollment data, while the occluded images (e.g. Fig. 4(e) and (f)) from the same person are used for testing. The similarity thresholds are learned in the same way as that in experiment 2 described in the previous section. Fig. 8 depicts the results as a function of the rank and cumulative match score, where our methods and the WLPS approach [16] are closely compared. It can be found

TABLE II

THE DETAILED RESULTS FOR EACH OF THE EXPRESSIONS THAT ARE LEFT OUT FOR TESTING ON THE AR DATABASE(%).

	Session1				Session2				
Algorithm	Neurtral	Smile	Angry	Scream	Neurtral	Smile	Angry	Scream	Average
NNC	94.0	100.0	98.0	82.0 *	65.0 *	71.0 *	78.0	44.0 *	79.0 *
GKLDA	**99.0**	97.0	**99.0**	90.0	82.0	79.0	83.0	56.0	85.6
SRC	98.0	100.0	97.0	93.0	77.0 *	82.0	81.0	60.0	86.0
LNMF	97.0	97.0	90.0	70.0 *	70.0 *	71.0 *	68.0 *	37.0 *	77.1 *
ICA	97.0	97.0	92.0	72.0 *	84.0	77.0 *	74.0 *	43.0 *	79.5 *
WLPS [18]	96.0	97.0	90.0	83.0 *	74.0 *	77.0 *	75.0 *	62.0	81.8 *
cPD	99.0	100.0	97.0	90.0	82.0	81.0	83.0	58.0	86.3
dPD	97.0	100.0	97.0	93.0	83.0	85.0	84.0	59.0	87.3
PD	98.0	**100.0**	97.0	**93.0**	**86.0**	**88.0**	**86.0**	**63.0**	**88.9**

* The asterisks indicate a statistically significant difference between the compared method and the PD method at a significance level of 0.05.

that our methods consistently achieve better performance than WLPS in both occlusion cases. In particular, on the sunglasses test, our PD method armed with the discriminant similarity threshold achieves a top 1 matching rate as high as 98.0%, significantly outperforming WLPS ($Z = 2.9$); While on the scarf tests, although the performance difference between the two is not statistically significant, the PD method achieves 10.0% higher performance than WLPS concerning the top 1 matching rate.

The comparison among our methods and some other methods under the two kinds of occlusions has been summarized in Table I. Although the performance of most of the compared algorithms under occlusions is in general poor, our methods yield superior performance to those methods. This is because of the unique ability of our methods of automatically excluding those sub-blocks in occlusions from matching, thus reducing the influence of those “useless” or even “harmful” sub-blocks as much as possible. In fact, the sub-blocks in occlusions generally would produce such a large deformation from the “normal” sub-blocks that they are beyond the acceptable similarity thresholds.

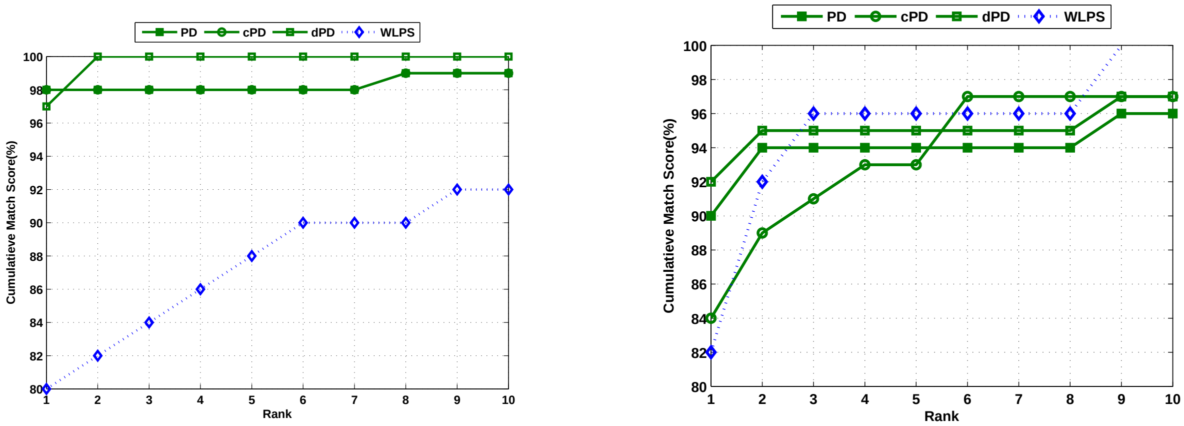
Nevertheless, it is worth mentioning that, as revealed in Table I, different algorithms exhibit different behaviors with respect to the ways of occlusion. For example, the LFA [22] method is more robust to lower face occlusion than upper-face occlusion, while LNMF [13], ICA [2] and our methods show the opposite behavior. This is somewhat surprising since it was believed that the upper face carries more discriminant information than the lower face. We believe that this is mainly because different approaches use different local features for face representation. For example, the LFA method heavily utilizes the geometric features near some pre-defined salient facial regions (e.g., eyes), so, occlusions in those regions may cause substantial degradation in the recognition performance. On the other hand, in LNMF, ICA and our methods, the appearance-based features are used, thus they are more insensitive to facial regions with high geometric complexity (such as eyes). The above discussion suggests that it is meaningful to further study the occlusion problem with hybrid local features.

E. Variations in both Partial Occlusion and Expressions

To further verify the robustness performance of our methods against both partial occlusions and expressions, we conducted a series of experiments on the FERET database [24]. The FERET database consists of more than 13,000 facial images corresponding to more than 1,500 subjects. The diversity of the database is across gender, ethnicity, and age. The data set used in our experiments consists of 2,400 FERET frontal face images corresponding to 1,200 subjects, with two images per subject (the index of these 2,400 images can be found at http://parnec.nuaa.edu.cn/dataset/feret.htm). The data set is further divided into three sets, i.e., the training set, the gallery set and the probe set, respectively. According to the FERET testing protocol [24], there is no overlap between these three sets. In particular, the training set consists of 480 images from 240 subjects, and for the two images from the left 960 subjects (i.e., except those 240 subjects used in the training set), one is put into the gallery set and the other into the probe set, thus one obtains 960 images each for the gallery and the probe set. Please refer to Section V-A for the details of the registration and preprocessing procedure performed on those images.

We simulate partial occlusions in each test image by using a black patch of size $p \times p$ with $p \in \{10, 20, \ldots, 50\}$ at a random location, see Fig. 9 for examples. Since the image size is only 80×60 pixels, the recognition task will become more and more challenging with the increasing of patch size. At each occlusion level, say 20×20, the simulation is repeated 10 times for each probe and we present the average recognition rate based on this. As before, our methods are compared against several state of the art methods designed to be robust against occlusions, including ICA [2], LNMF [13] ,partitioned SRC [37](with tuned block size 4×4) and the Kernel LDA with Local Gabor features (GKLDA, [15]). Notice that the GKLDA method has achieved state of the art performance on the FRGC database [15].

The results are shown in Fig.10. The proposed PD algorithm significantly outperforms the others, for all levels of occlusions. In particular, under the extreme occlusion with patch size 50×50 (which means 52.1% image is occluded - a challenging recognition task even for humans, see last row in Fig.9), the proposed PD algorithm can still reach an average rank 1 recognition rate of 78.5%, compared to SRC’s 47.8% and KLDA’s 11.9%. The figure also shows that the overall stableness performance of our method against occlusion is also superior to that of SRC and GKLDA. In fact, for SRC and GKLDA, the standard variance of the performance over different occlusion patch sizes is respectively 15.1% and



(a)

(b)

Fig. 8. Results on partial occlusions on the AR Databases. (a) sunglasses and (b) scarf.

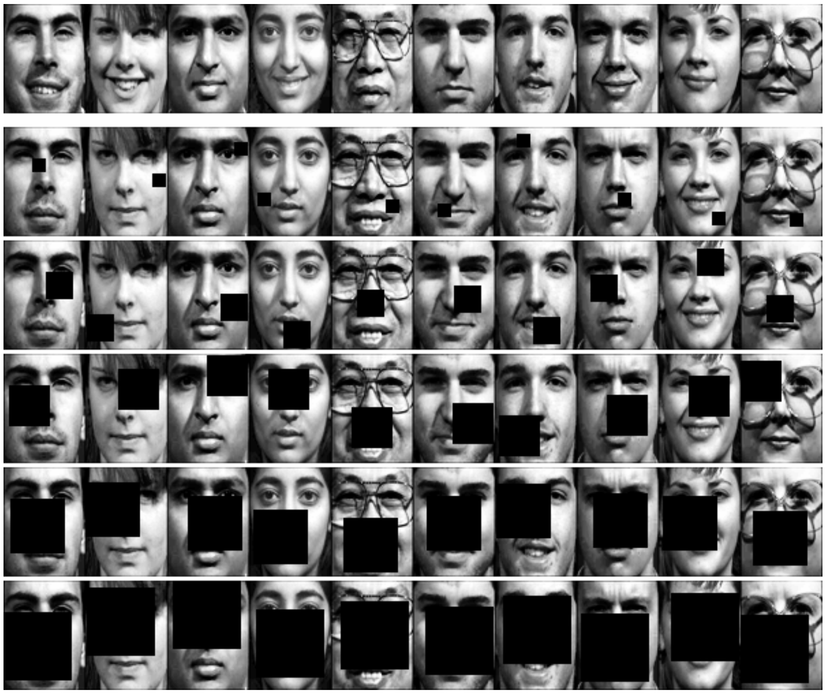


Fig. 9. Examples of FERET face images with simulated occlusions. The top row shows 10 gallery images from 10 different subjects, while in the following five rows each displays the corresponding test images with random occluding patches of sizes (from top to bottom) 10x10, 20x20, ..., 50x50).

30.6%, while that of the PD algorithm is only 4.1%.

It is worthy mentioning that these results also add useful experimental evidence to the long time debate between the local method and global method in face recognition community [43]. In our opinion, since the manner of facial appearance changing due to local variations and global ones are different, they may be best addressed using different strategy: global models usually rely on linear/nonlinear transformations (e.g. KLDA [15], ICA [2],LNMF [13]) or sparsely combinations (e.g., global SRC [37]) of the whole vectors of face images to fit the data, and thus tend to be more robust against holistic appearance variations due to lighting, ageing or small amount of local deformation, but the underlying holistic basis are also less likely to well fit the image data deformed by large local variations caused by partial occlusions or disguise. As can be seen from Fig.10, although GKLDA is a good competitor to handle lighting, ageing effect and other holistic variations, its performance decreases significantly when about 18.8% (i.e.,30 × 30 patch size) face images is occluded. In these cases, spatially local model such as ours would be a better choice.

F. Other facial distortions

To gain some insight on the robustness of our methods against slight variations of pose angle and alignment, we test our approaches on the ORL face database. The experiments follow the testing protocol used in several previous works [12], [14]. That is, five images of each person are randomly selected for the gallery set and the other five images for the probe set.

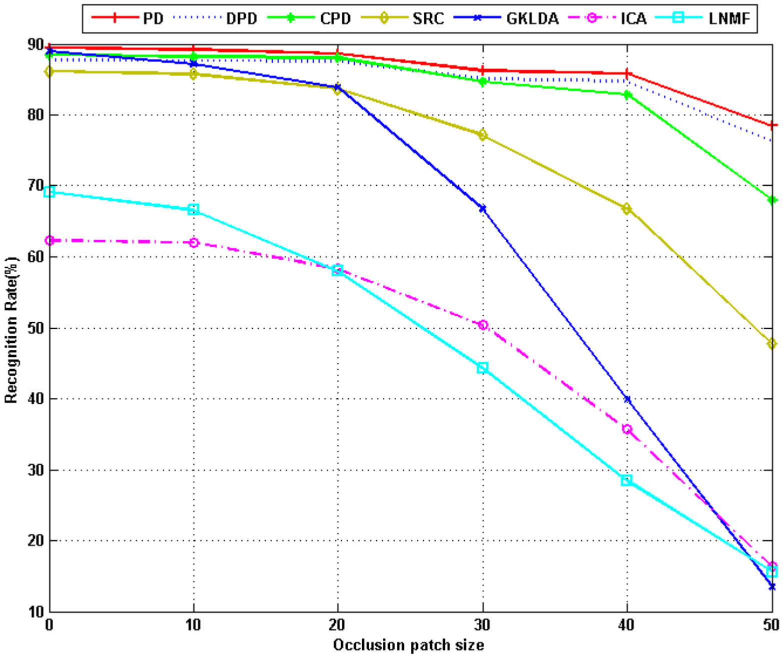


Fig. 10. Comparative Performance between various algorithms on the FERET Databases with varying level of random occlusion (in 10x10, 20x20, ..., 50x50 of occluding patches), with 200, 218 basis components for the ICA and LNMF respectively, and Gaussian kernel for GKLDA (the values of kernel parameters are optimized with cross-validation) .

We adopt the same definition of average correct matching rate, R_{ave}, used in [12], which is given by

$$R_{ave} = \frac{\sum_{i=1}^{q} n_{corr}^{i}}{q n_{tot}} \tag{18}$$

where q is the number of experimental runs, n_{corr}^{i} is the number of correct classifications for the ith run, and n_{tot} is the number of total testing patterns of each run. Moreover, following [12], we also test our methods with varying training sizes per subject. The results of 50 runs of the experiments are summarized in Table III. As mentioned in Section V-A, we don't make any preprocessing on the images of the ORL databases. Table III reveals that our methods are robust against slight imprecise alignment.

G. Specific Issues Concerning the Partial Distances

Some specific issues involved in the proposed methods are studied in this section.

1) The Effect of Different Sub-block Sizes: To investigate the effect of different sub-block size on the performance, we repeat the previous experiments conducted with various sub-block sizes from small (2 × 2) to large (10 × 10). The rank 1 matching score as a function of the sub-block size are plotted in Fig. 11. Notice that only the results yielded by the PD method are shown in this figure because the results of the *c*PD and *d*PD are very similar to those of PD.

Fig. 11 reveals that the performance of the proposed algorithm is quite robust against different sub-block sizes but smaller sub-block size tends to obtain better performance than larger one due to the loss of spatial information with increasing block size. On the other hand, smaller block size results in more blocks and larger memory requirement (see the next section for more on this), and we recommend a block size of 4×4 pixels in practice.

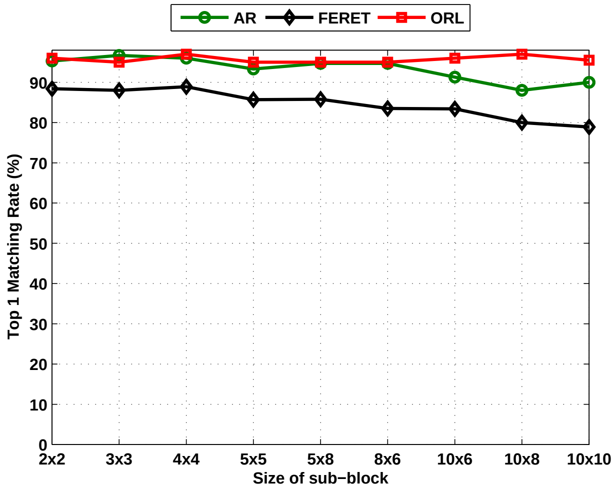


Fig. 11. Recognition rate as a function of the size of sub-blocks.

2) On the Computational Complexity: In the proposed methods, most processing time cost goes to the computation of local similarity matrix D (see Eq. 2). Let the size of training set be N. Suppose that each face in dim_a dimension is partitioned into K sub-blocks with dimensionality dim_b. Then the computational complexity for a similarity matrix is $O(dim_b KN)$. Note that this is equal to the computational complexity of the standard nearest neighbor rule, i.e. $O(dim_a N)$, since $dim_a = dim_b K$. As mentioned before, the matrix is computed in the SOM topological space, thus the computational cost is actually reduced to $O(2KN)$. On our machine with 800MHz processor and 512MB RAM, the

TABLE III
EXPERIMENTAL RESULTS WITH VARYING TRAINING SIZES ON THE ORL DATABASE(%).

Training	PCA	SOM+CN	NFL	NMF	LNMF	The proposed method		
Size	[12]	[12]	[14]	[13]	[13]	cPD	dPD	PD
3	81.8	88.2	N/A	N/A	N/A	91.8	91.7	**92.5**
5	89.5	96.5	96.9	91.0	94.0	96.9	97.0	**97.3**

SOM-mapping scheme generally runs about 10 times or more faster than that without embedding, demonstrating a significant improvement in efficiency.

The SOM-based representation could also reduce the space requirement for a face database from $O(dim_a N)$ to $O(2KN)$. Nevertheless, the space complexity for a similarity matrix remains $O(KN)$. This problem can be alleviated by using techniques such as editing or condensing [21], by eliminating "useless" prototypes. It is one of the future works to implement the algorithm with lower space complexity.

VI. Conclusions

The research reported here extends a preliminary research [30] which advocates using non-metric distance in face recognition. Various aspects are examined here, with particular emphasis on how to detect and capture the prominent partial similarity among face images while exclude unreliable and unimportant features. For that purpose, we present two similarity threshold setting strategies to distinguish the most useful image patches from those less useful ones, based on a local similarity representation of face images. The first strategy is based on the golden section rule which empirically excludes less useful portions from recognition, while the other is based on the maximum marginal criterion, allowing one to learn the optimal intra-personal partial similarity for each class. The feasibility and effectiveness of these two strategies is verified by a series of experiments on several well-known face databases. It is found that partial similarity matching performs better than several other methods in handling expression variations, partial occlusions and other local distortions.

It is worthy mentioning that besides face recognition, the proposed large-margin-based similarity threshold learning technique may also find applications in such areas as semi-supervised clustering, image segmentation and outlier detection. The learning of discriminative partial similarity may also be used as a building block for other more sophisticating methods, such as for constructing a robust kernel matrix.

Acknowledgment

We want to thank the anonymous reviewers for their valuable comments and suggestions and Dr. Aleix M. Martinez for providing the AR database. Portions of the research in this paper use the FERET database of facial images collected under the FERET program. This research was supported by the National Science Foundation of China (60773060, 60635030, 60721002), the Jiangsu Science Foundation (BK2006187,BK2008018), the National High Technology Research and Development Program of China (2007AA01Z169), the Jiangsu 333 High-Level Talent Cultivation Program, the Innovation Foundation of NUAA (Y0603-042) and the Project-sponsored by SRF for ROCS, SEM.

Xiaoyang Tan received his B.S. and M.S. degree in computer applications from Nanjing University of Aeronautics and Astronautics (NUAA) in 1993 and 1996, respectively. Then he worked at NUAA in June 1996 as an assistant lecturer. He received a Ph.D. degree from Department of Computer Science and Technology of Nanjing University, China, in 2005. From Sept.2006 to OCT.2007, he worked as a postdoctoral researcher in the LEAR (Learning and Recognition in Vision) team at INRIA Rhone-Alpes in Grenoble, France. His research interests are in face recognition, machine learning, pattern recognition, and computer vision.In these fields, he has authored or coauthored over 20 scientific papers.

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 NUAA 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-SM'06) received the BSc, MSc and PhD degrees in computer science from Nanjing University, China, in 1996, 1998 and 2000, respectively, all with the highest honors. He joined the Department of Computer Science & Technology, Nanjing University, as a Lecturer in 2001, and is currently Cheung Kong Professor and Director of the LAMDA group. His research interests are in artificial intelligence, machine learning, data mining, information retrieval, pattern recognition, evolutionary computation and neural computation. In these areas, he has published over 60 papers in leading international journals or conference proceedings. Dr. Zhou has won various awards/honors including the National Science & Technology Award for Young Scholars of China (2006), the Award of National Science Fund for Distinguished Young Scholars of China (2003), the National Excellent Doctoral Dissertation Award of China (2003), the Microsoft Young Professorship Award (2006), etc. He is associate editor-in-chief of *Chinese Science Bulletin*, associate editor of *IEEE Transactions on Knowledge and Data Engineering*, and on the editorial boards of journals including *Artificial Intelligence in Medicine*, *Intelligent Data Analysis*, *Knowledge and Information Systems*, *Science in China*, etc. He is/was a steering committee member of PAKDD and PRICAI, program committee chair/co-chair of PAKDD'07 and PRICAI'08, vice chair or area chair of IEEE ICDM'06, IEEE ICDM'08, SIAM DM'09, etc., program committee member of various international conferences including AAAI, ICML, ECML, ACM SIGKDD, IEEE ICDM, SIAM DM, ACM Multimedia, etc., and general chair/co-chair or program committee chair/co-chair of a dozen of native conferences in China. He is a senior member of the China Computer Federation (CCF) and the vice chair of the CCF Artificial Intelligence & Pattern Recognition Society, an executive committee member of the Chinese Association of Artificial Intelligence (CAAI) and the chair of the CAAI Machine Learning Society, and the chair of the IEEE Computer Society Nanjing Chapter.

Jun Liu received his B.S. degree from Nantong Institute of Technology (now Nantong University) in 2002, and his Ph.D. degree from NUAA in November, 2007. He joined the Department of Computer Science & Engineering, NUAA, as a Lecturer in 2007. He is currently a postdoc in the Biodesign Institute of Arizona State University. His research interests include Pattern Recognition and Computer vision, and he has authored or coauthored over 10 scientific papers.

REFERENCES

[1] B. H. Aharon, T. Hertz, N. Shental, and D. Weinshall, "Learning a Mahalanobis metric with side information," *Journal of Machine Learning Research*, vol. 6, pp. 937–965, 2005.

[2] M. Bartlett, J. Movellan, and T. Sejnowski, "Face recognition by independent component analysis," *IEEE Transactions on Neural Networks*, vol. 13, no. 6, pp. 1450–1464, 2002.

[3] P. Belhumeur, J. Hespanha, and D. Kriegman, "Eigenfaces vs. Fisherfaces: Recognition using class specific linear projection," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 19, no. 7, pp. 711–720, 1997.

[4] P. J. Flynn, K. W. Bowyer, and P. J. Phillips, "Assessment of time dependency in face recognition: An initial study," in *Lecture Notes in Computer Science 2688*, J. Kittler and M. S. Nixon, Eds. Berlin: Springer, 2003, pp. 44–51.

[5] R. Gross, J. Cohn, and J. Shi, "Quo vadis face recognition," in *Working Notes of the 3rd Workshop on Empirical Evaluation Methods in Computer Vision*, Kauai, HI, 2001.

[6] X. F. He, X. C. Yan, Y. Hu, P. Niyogi, and H. J. Zhang, "Face recognition using Laplacianfaces," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 27, no. 3, pp. 328–340, 2005.

[7] G. R. Hjaltason and H. Samet, "Properties of embedding methods for similarity searching in metric spaces," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 25, no. 5, pp. 530–549, 2003.

[8] Y. Ivanov, B. Heisele, and T. Serre, "Using component features for face recognition," *Automatic Face and Gesture Recognition, 2004. Proceedings. Sixth IEEE International Conference on*, pp. 421–426, May 2004.

[9] D. W. Jacobs, D. Weinshall, and Y. Gdalyahu, "Classification with non-metric distances: Image retrieval and class representation," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 22, no. 6, pp. 583–600, 2000.

[10] J. Kittler, M. Hatef, R. Duin, and J. Matas, "On combining classifiers," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 20, no. 3, pp. 226–239, 1998.

[11] T. Kohonen, *Self-Organizing Map*, 2nd ed. Berlin: Springer, 1997.

[12] S. Lawrence, C. Lee Giles, A. Tsoi, and A. Back, "Face recognition: A convolutional neural-network approach," *IEEE Transactions on Neural Networks*, vol. 8, no. 1, pp. 98–113, 1997.

[13] S. Z. Li, X. W. Hou, and H. J. Zhang, "Learning spatially localized, parts-based representation," in *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, Kauai, HI, 2001, pp. 207–212.

[14] S. Z. Li and J.Lu, "Face recognition using the nearest feature line method," *IEEE Transactions on Neural Networks*, vol. 10, no. 2, pp. 439–443, 1999.

[15] C. Liu, "Capitalize on dimensionality increasing techniques for improving face recognition grand challenge performance," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 28, no. 5, pp. 725–737, 2006.

[16] M. Martinez, "Recognizing imprecisely localized, partially occluded, and expression variant faces from a single sample per class," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 24, no. 6, pp. 748–763, 2002.

[17] M. Martinez and R. Benavente, "The AR face database," CVC, Tech. Rep. 24, 1998.

[18] M. Martinez and Y. Zhang, "Subset modeling of face localization error, occlusion, and expression," in *Face Processing: Advanced Modeling and Methods*, R. Chellappa and W. Zhao, Eds. Academic Press, 2005.

[19] B. Moghaddam and A. Pentland, "Probabilistic visual learning for object representation," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 19, no. 7, pp. 696–710, 1997.

[20] B.-G. Park, K.-M. Lee, and S.-U. Lee, "Face recognition using face-ARG matching," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 27, no. 12, pp. 1982–1988, 2005.

[21] E. Pekalska, R. P. W. Duin, and P. Paclik, "Prototype selection for dissimilarity-based classification," *Pattern Recognition*, vol. 39, no. 2, pp. 189–208, 2006.

[22] P. Penev and J. Atick, "Local feature analysis: A general statistical theory for object respresentaion," *Network: Computation in Neural Systems*, vol. 7, no. 3, pp. 477–500, 1996.

[23] P. J. Phillips, P. J. Flynn, W. T. Scruggs, K. W. Bowyer, J. Chang, K. Hoffman, J. Marques, J. Min, and W. J. Worek, "Overview of the face recognition grand challenge," in *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, San Diego, CA, 2005, pp. 947–954.

[24] P. J. Phillips, H. Moon, S. A. Rizvi, and P. J. Rauss, "The FERET evaluation methodology for face-recognition algorithms," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 22, no. 10, pp. 1090–1104, 2000.

[25] A. N. R. Singh, M. Vatsa, "Face recognition with disguise and single gallery images," *Image Vision Comput*, p. in press, 2008.

[26] S. T. Roweis and L. K. Saul, "Nonlinear dimensionality reduction by local linear embedding," *Science*, vol. 290, no. 5500, pp. 2323–2326, 2000.

[27] S. Santini and R. Jain, "Similarity measures," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 21, no. 9, pp. 871–883, 1999.

[28] L. Sirovich and M. Kirby, "Low dimensional procedure for the characterization of human faces," *Journal of Optical Society of America*, vol. 4, no. 3, pp. 519–524, 1987.

[29] A. P. Stakhov, *Codes of the Gold Proportion*. Moscow, Russia: Radio and Communication Publishing House, 1984.

[30] X. Tan, S. Chen, Z.-H. Zhou, and J.Liu, "Learning non-metric partial similarity based on maximal margin criterion," in *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, New York, NY, 2006, pp. 138–145.

[31] X. Tan, S. Chen, Z.-H. Zhou, and F. Zhang, "Recognizing partially occluded, expression variant faces from single training image per person with SOM and soft kNN ensemble," *IEEE Transactions on Neural Networks*, vol. 16, no. 4, pp. 875–886, 2005.

[32] ——, "Face recognition from a single image per person: A survey," *Pattern Recognition*, vol. 39, no. 9, pp. 1725–1745, 2006.

[33] T.Cox and M.Cox, *Multidimensional Scaling*, Chapman and Hall, London, 1994.

[34] A. Timo, H. Abdenour, and P. Matti, "Face recognition with local binary patterns," in *Proceedings of the 8th European Conference on Computer Vision*, Prague, Czech, 2004, pp. 469–481.

[35] M. Turk and A. Pentland, "Eigenfaces for recognition," *Journal of Cognitive Neurscience*, vol. 3, no. 1, pp. 71–86, 1991.

[36] P. Viola and M. Jones, "Robust real-time object detection," *International Journal of Computer Vision*, vol. 57, no. 2, pp. 137–154, 2004.

[37] J. Wright, A. Yang, A. Ganesh, S. S., and Y. Ma, "Robust face recognition via sparse representation," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, to appear.

[38] www.wikipedia.com.

[39] E. P. Xing, A. Y. Ng, M. I. Jordan, and S. Russell, "Distance metric learning with application to clustering with side information," in *Advances in Neural Information Processing Systems 15*, S. Becker, S. Thrun, and K. Obermayer, Eds. Cambridge, MA: MIT Press, 2003, pp. 505–512.

[40] W. Yambor, B. Draper, and R. Beveridge, "Analyzing PCA-based face recognition algorithms: Eigenvector selection and distance measures," in *Working Notes of the 2nd Workshop on Empirical Evaluation Methods in Computer Vision*, Dublin, Ireland, 2000.

[41] J. Yang, A. F. Frangi, J.-Y. Yang, D. Zhang, and Z. Jin, "KPCA plus LDA: A complete kernel fisher discriminant framework for feature extraction and recognition," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 27, no. 2, pp. 230–244, 2005.

[42] W. Zhang, S. Shan, W. Gao, and H. Zhang, "Local Gabor Binary Pattern Histogram Sequence (LGBPHS): A novel non-statistical model for face representation and recognition," in *Proceedings of the IEEE International Conference on Computer Vision*, Beijing, China, 2005, pp. 786–791.

[43] W. Zhao, R. Chellappa, P. J. Phillips, and A. Rosenfeld, "Face recognition: A literature survey," *ACM Computing Survey*, vol. 34, no. 4, pp. 399–485, 2003.