

Learning Non-Metric Partial Similarity Based on Maximal Margin Criterion

Xiaoyang Tan^{1,2} Songcan Chen¹ Zhi-Hua Zhou² Jun Liu¹

¹Department of Computer Science and Engineering
Nanjing University of Aeronautics & Astronautics, Nanjing 210016, China

²National Laboratory for Novel Software Technology
Nanjing University, Nanjing 210093, China

{x.tan,s.chen,j.liu}@nuaa.edu.cn zhouzh@nju.edu.cn

Abstract

The performance of many computer vision and machine learning algorithms critically depends on the quality of the similarity measure defined over the feature space. Previous works usually utilize metric distances which are often epistemologically different from the perceptual distance of human beings. In this paper, a novel non-metric partial similarity measure is introduced, which is born to automatically capture the prominent partial similarity between two images while ignoring the confusing unimportant dissimilarity. This measure is potentially useful in face recognition since it can help identify the inherent intra-personal similarity and thus reducing the influence caused by large variations such as expression and occlusions. Moreover, to make this method practical, this paper proposes an automatic and class-dependent similarity threshold setting mechanism based on the maximal margin criterion, and uses a Self-Organization Map-based embedding technique to alleviate the computational problem. Experimental results show the feasibility and effectiveness of the proposed method.

1 Introduction

Similarity measure has attracted much attention of researchers from diverse areas such as computer vision, machine learning and pattern recognition during the past few years [6, 14, 16]. Actually, the underlying similarity measure has great impact on the performance of many learning algorithms such as clustering algorithms and nearest neighbor classifiers. Therefore, how to choose a “good” similarity measure is one of the key concerns of these algorithms.

This paper focuses on the problem of modelling similarity for direct face image matching. Face image matching algorithms seek to identify a target image from a large face database with maximal similarity to a given probe face im-

age [21]. The success of such a task critically depends on the quality of the similarity measure defined over the image space. Euclidean distance is the most widely used similarity measure. However, face images are generally assumed to span a low-dimensional nonlinear manifold in a high-dimensional space [4] and therefore, to obtain a good generalization performance, the prior defined Euclidean distance is not always appropriate. This is particular true if the given face database contains complex intra-personal variations caused by illuminations, occlusions and expressions.

In fact, metric distances such as Euclidean distance is subject to the rigid constraints of metric axioms (i.e., self-similarity, symmetry, triangle inequality), and several recent studies have shown that these metric axioms are epistemologically invalid for perceptual distance of human beings [14, 16] and not so suitable for robust pattern recognition [6]. As indicated by Jacobs et al. [6], the changeful face images cannot be matched into a metric feature space without large distortions in the distances between them. Therefore, non-metric distance functions seem to be a natural choice for robust face image matching, which is the main concern of this paper.

Non-metric distances can be developed by part-based methods, which uses image patches and has the inherent advantage of capturing local statistical relationships among specific pixels in an image. Popular part-based methods include NMF (Non-negative Matrix Factorization) [8], ICA (Independence Component Analysis) [2], Local Probabilistic Subspace [9, 10], etc. These methods are shown to be successful in dealing with face images with complex intra-personal variations.

In this paper, we present a novel part-based non-metric distance learning method. The idea can be best illustrated by Figure 1, where many observers will find that both the person and the horse are similar to the centaure, but the person and the horse are not similar to each other. A reasonable explanation is that when comparing two images, we

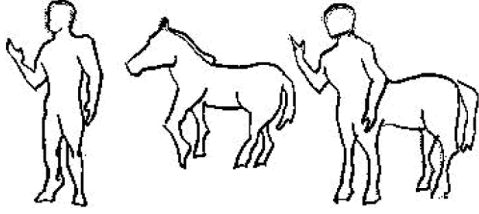


Figure 1. An illustration of the intuition of partial similarity judgments.

human beings tend to “focus on the portions that are very similar and are willing to pay less attention to regions of great dissimilarity” [6]. In other words, in contrast to dissimilar portions, similar portions are more important and play the dominant role in the process of image matching. This observation implies that the similar portions may carry highly discriminant information for robust image comparison. Capturing such information, therefore, is of great importance to classification under high-dimensional settings.

In light of these observations, a goal of this paper is to design a mechanism to support the partial similarity measurement. Based on a local dissimilarity representation of face images, we propose a partial similarity measure that can automatically capture the significant partial similarity between two face images while ignoring the unimportant dissimilarity. By this way, we can evaluate the perceptual distance between face images more accurately, and lower the influence of appearance variations presented in images.

By *partial similarity*, we mean that the similar regions of the face images will be accentuated in the process of face matching. Thus, how to define *similar regions* becomes the critical issue of the proposed method. This paper addresses this problem with a maximal margin-based learning criterion, making the similarity threshold setting become automatic and adaptive in nature.

In addition, to improve the computational efficiency, an SOM (Self-Organization Map [7])-based embedding strategy is employed. Experimental results show that the SOM can not only embed the facial portions into a low-dimensional space faithfully, but also filter out the possible noise at the same time, resulting in an encouraging enhancement to the proposed method.

The rest of this paper is organized as follows. Section 2 briefly reviews some related works. Section 3 describes the proposed partial distance measure. Section 4 introduces the SOM embedding. Section 5 describes how to learn the similarity threshold with the maximal margin criterion. Section 6 reports on the experiments. Finally, Section 6 concludes.

2 Related Works

Non-metric distances have attracted attention from researchers in the field of face recognition. Bayesian method

[12] and kernel-based method [20] are two representative works along this direction. The former uses a probabilistic measure of similarity based primarily on a Bayesian analysis of image differences, while the latter seeks to find a non-linear transformation of the similarity between two images in the input space such that the class boundaries more likely become linear in the transformed space.

In the field of machine learning, *metric learning* has become very hot during the past few years [1, 19]. Most works seek to formulate the metric learning problem as some kind of mathematical programming problem, with the hope of finding the optimal parameters that minimize some cost function. Xing’s method [19] and the Relevant Component Analysis (RCA) distance [1] are representative works. Both employ extra information (e.g., equivalence constraints) to learn the similarity between samples. Note that their optimal criteria are based on the whole training set, which forces a relatively strong constraint on the data set, thus increasing the complexity of the optimization problem.

All the aforementioned methods do not take the spatial structure of image data into account. A most recent work which explicitly considers this problem is the so-called Image Euclidean Distance (IMED) [18], where the spatial information of pixels in an image are exploited. Our work differs from the above work in both the ways of face image representation and the optimization method taken.

3 The Proposed Partial Distance Measure

The overall framework of the proposed method is shown in Figure 2. To enable partial similarity capturing, both the training face images and the probe image are localized into sub-blocks in the same way at first. Then, all the obtained sub-blocks are embedded in a pre-trained SOM topological space, where a nearest neighbor search is executed using the proposed partial distance measure, and the training face image reporting the smallest distance is selected to give the final identity.

Among all the possible definitions of local facial region, perhaps the simplest is the one that defines local facial re-

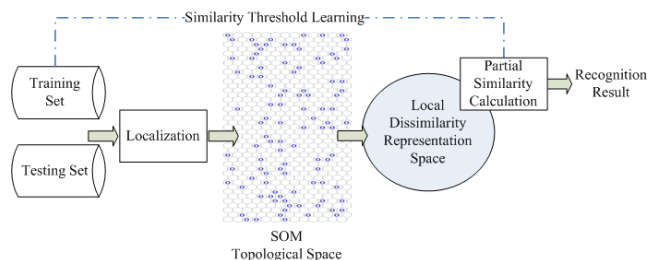


Figure 2. Overall framework of the proposed method.

gion as a rectangle or sub-block in the image. In particular, each face image is partitioned into K ($= \dim_a / \dim_b$) non-overlapping sub-blocks with equal size, where \dim_a and \dim_b are the dimensionalities of the whole image and each sub-block, respectively. For simplicity and efficiency, each sub-block is represented as a local feature vector (LFV) by concatenating the pixels of the sub-block. Such a gray-level-based local feature representation has been proven to be helpful in both face detection [17] and face recognition [15].

3.1 Local Pairwise Dissimilarity Matrix

After defining the sub-blocks of face image, we construct the local pairwise dissimilarity matrix D between a probe face \mathbf{q} and every face \mathbf{x}_i in the training set T . Denote the sub-blocks of \mathbf{q} and \mathbf{x}_i as $\{\mathbf{q}^k\}_{k=1}^K$ and $\{\mathbf{x}_i^k\}_{k=1}^K$, respectively. Thus, the component at the k -th row and i -th column of D is $d_{ki} = d(\mathbf{q}^k, \mathbf{x}_i^k)$, which is the local pairwise distance between the k -th sub-blocks of the probe face and the i -th training face. Although from the holistic view, face images span a non-metric manifold, it is usually assumed that the neighborhood of local patches of face images can be well approximated by metric distance. Hence we have:

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

where $(\|\cdot\|_p)$ is the L_p -norm defined on the LFVs \mathbf{q}^k and \mathbf{x}_i^k with $p \geq 1$). Then the dissimilarity matrix is:

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

where N is the total number of training faces in T . This matrix plays an important role in our method. Actually it contains all the information needed for the subsequent recognition task. In terms of the similarity approaches, this matrix can be regarded a local dissimilarity representation of the data.

Based on the above notations, we define the global distance $d(\mathbf{q}, \mathbf{x}_i)$ between the probe face \mathbf{q} and the training image \mathbf{x}_i as:

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

That is, the global distance is approximated by the sum of the local pairwise distances. Here for simplicity and without loss of generality, we use L_1 -norm to calculate the local pairwise distances.

3.2 Partial Distance Measure

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

$$S = \{k | d_{ki} \leq \tau, k = 1, \dots, K\} \quad (4)$$

$$F = \{k | d_{ki} > \tau, k = 1, \dots, K\} \quad (5)$$

S and F are called the *similar* subset and *dissimilar* subset, respectively. Note that by the definition, two sub-blocks can be regarded as similar even though they have relatively big distance (e.g., with d_{ki} close to τ), which could be helpful for improving the generalization ability. Based on these notations, we can rewrite Eq. 3 to:

$$d(\mathbf{q}, \mathbf{x}_i) = \sum_{k=1}^K d_{ki} = \sum_{k \in S} d_{ki} + \sum_{k \in F} d_{ki} \quad (6)$$

That is, the global distance between two face images is equal to the linear sums of the local pairwise distances between similar portions and dissimilar portions. Furthermore, Eq. 6 can be generalized as:

$$d'(\mathbf{q}, \mathbf{x}_i, \tau) = \beta \sum_{k \in S} d_{ki} + (1 - \beta) \sum_{k \in F} d_{ki} \quad (7)$$

where $\beta \in [0, 1]$ is a parameter controlling the weight of similar and dissimilar portions in similarity measuring. β can be set based on statistics of the similar and dissimilar portions. Let $|S|$ and $|F|$ denote the number of similar and dissimilar sub-blocks, respectively. Then, β can be set as:

$$\beta = \min(1, \frac{|S|}{|F|}) = \min(1, \frac{|S|}{K - |F|}) \quad (8)$$

Clearly, β is in $[0, 1]$ and as β approaching 1, similar portions will play a more important role in the global distance calculation. Indeed, the value of β is solely controlled by the number of similar portions $|S|$, which is in turn determined by the underlying threshold.

In our implementation, for simplicity, the pairwise sub-block distance is discretized into an integer as follows:

$$I(d_{ki}) = \begin{cases} 1, & d_{ki} \leq \tau \\ 0, & d_{ki} > \tau \end{cases} \quad (9)$$

Accordingly, Eq. 7 is rewritten to:

$$\begin{aligned} d_p(\mathbf{q}, \mathbf{x}_i, \tau) &= \beta \sum_{d_{ki} \leq \tau} I(d_{ki}) + (1 - \beta) \sum_{d_{ki} > \tau} I(d_{ki}) \\ &= \beta \sum_{d_{ki} \leq \tau} I(d_{ki}) \end{aligned} \quad (10)$$

That is, the distance between two faces completely depends on the weighted number of similar sub-blocks. Hence

the name *partial distance* (PD). Note that the weighting coefficient β implicitly takes the possible influence of the dissimilar portions into account.

Finally, the subject identification can be executed according to the nearest neighbor rule using the partial distance defined above:

$$label(\mathbf{q}) = arg \min_{i=1 \dots N} (d_p(\mathbf{q}, \mathbf{x}_i, \tau)) \quad (11)$$

where $label(\mathbf{q})$ is the label of an unknown face \mathbf{q} .

3.3 Properties of the Proposed Measure

The partial distance defined above has some interesting properties that are desired in face image matching. First, it automatically selects the most similar portions between two faces for comparison, which actually makes the complex intra-personal distribution being more compact.

Second, by definition, a distance measure is a metric distance if it satisfies the metric axioms, i.e., non-negativity, self-similarity, symmetry, and the triangle inequality [14]. If any one of the above conditions is violated, the concerned distance measure is called non-metric distance. It is obvious that both the non-negativity and symmetry are satisfied by the proposed partial distance. However, it is not transmissive, i.e., it violates the triangle inequality (transitivity should be followed from the triangle inequality [6]). This occurs mainly because different sub-blocks can make contribution in different comparisons. As for the centaure example shown in Fig. 1, similar sub-blocks between the person and centaure and these between the horse and centaure are different. This suggests that the triangle inequality may violate the perceptual similarity of human beings.

Moreover, the proposed partial distance doesn't satisfy the self-similarity axiom as well. That is, two face images may have *zero* partial distance even if they are not identical. This looks surprising at first glance since it implies that given a probe face \mathbf{q} , its nearest neighbor may not necessarily be itself. But note that in reality, the requirement that only identical objects would yield a zero distance may be too strong. Actually, in face recognition, there are rarely two absolutely identical face images even from the same person. Therefore, it seems that allowing two face images with slight deformation to be recognized as the same could be a better strategy, since this potentially increases the possibility of finding the correct matching for a given face. This property turns out to be a major difference between the proposed partial distance and any metric distance measures.

4 Embedding with SOM

It is evident that the direct calculation of pairwise distances in the input space is computationally expensive. A

feasible solution is to map, or embed the local facial vectors into a low-dimensional embedding space such that [5]: (1) the distances of the embedded vectors approximate the actual distances, and (2) the dissimilarity matrix computation can be performed in the "less expensive" embedding space.

In this paper, Self-Organizing Maps [7], one of the most efficient and effective techniques that can meet the above two requirements simultaneously, is adopted. More specifically, after localizing the images, an SOM network is trained and used to project all the LFVs onto a quantized lower-dimensional space, in which the dissimilarity matrix (Eq.2) is then calculated.

In the proposed method, most of the time cost goes to the computation of the local dissimilarity matrix D with complexity of $O(dim_b KN)$. Note that this is equal to the computational complexity of the standard nearest neighbor rule, i.e., $O(dim_a N)$, since $dim_a = dim_b K$. Due to the SOM embedding, the computational cost is actually reduced to $O(2KN)$. On a machine with 800MHz processor and 512MB RAM, after using the SOM-embedding scheme, the proposed method generally runs about 10 times or more faster than before.

5 Learning Similarity Threshold with the Maximal Margin Criterion

The proposed method involves an important parameter which needs to be adjusted, i.e., the similarity threshold τ (Eq.4). This threshold defines the minimal acceptable distance between two sub-blocks, and all the pairwise sub-blocks with distances below the threshold are considered to be similar to each other.

Intuitively, a good threshold should be class-dependent in nature, i.e., different thresholds should be used for different persons (classes). In order to ensure a low generalization error, the threshold should also assign a training face images to the correct class with high confidence, which is related to *margin*. However, in most large margin practice, such as RCA [1], the margin over the whole training set is optimized, which is a relatively strong constraint on the cost function, potentially increasing the complexity of the optimization problem. This paper does not require all the similar samples be clustered tightly. Instead, the margin of each class is optimized separately, one for each time. Such a 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 sample \mathbf{x}_i . The index set of the training samples belonging to the c -th class is $H_c = \{i | y_i = c, i = 1, \dots, N\}$, and the index set of the samples from other classes is $\bar{H}_c = \{i | y_i \neq$

$c, i = 1, \dots, N\}$. Then, the training set of the c -th class is denoted as $X_c = \{\mathbf{x}_i, i \in H_c\}$ with size $|H_c|$. Furthermore, denote the threshold of the c -th class as τ_c .

The optimization for τ_c involves a Leave-One-Out (LOO) validation strategy on X_c . First, fetch one sample $\mathbf{x}_i \in X_c$ as the validation sample, and all the remaining samples in the training set T as prototypes. Then, try to label the validation sample using the partial distance under a given threshold τ_c (with Eq. 11). Suppose the classification result is $LOO(\mathbf{x}_i, \tau_c)$, then the average margin of the c -th class can be defined as:

$$\bar{m}_c(X_c, \tau_c) = \frac{1}{|H_c|} \sum_{i \in H_c} \{1(LOO(\mathbf{x}_i, \tau_c) = y_i) [\min_{j \in H_c} d_{PD}(\mathbf{x}_i, \mathbf{x}_j, \tau_c) - \min_{\substack{j \in H_c \\ j \neq i}} d_{PD}(\mathbf{x}_i, \mathbf{x}_j, \tau_c)]\} \quad (12)$$

where $1(u)$ is the indicator function which takes 1 if u is true and 0 otherwise. Eq. 12 says that if a training sample $\mathbf{x}_i, i \in H_c$ is correctly classified in the Leave-One-Out validation, then the classification confidence is the margin between the nearest training sample of other classes (the first term) and the nearest prototype (except \mathbf{x}_i itself) of the c -th class (the second term). Clearly, only positive values of $\bar{m}_c(X_c, \tau_c)$ denote correct classifications, and the larger the value, the higher confidence the classification.

Eq. 12 is then used as the cost function to be optimized in the training phase and its output is the so-needed class-dependent similarity threshold τ_c^* .

Maximizing a margin function like Eq. 12 is generally difficult. Here a straightforward greedy search strategy is employed. That is, since the degree of similarity ranges in $[0, 1]$ (0 denotes totally dissimilar and 1 the highest similar), the range of $[0, 1]$ can be greedily searched to approximate the optimal solution. More specifically, the interval of $[0, 1]$ is discretized into $(h+1)$ candidate values, denoted by $\tau_{candidate} = [0, \frac{1}{h}, \frac{2}{h}, \dots, \frac{(h-1)}{h}, 1]$, where h is the parameter that controls the threshold searching granularity and is set to 100 in this paper. Then, the candidate value maximizing Eq. 12 is deemed as the best approximation to the optimal threshold τ_c^* .

Before that, however, the components of the dissimilarity matrix D (Eq. 2) should be turned into similarity in order to ensure that every component is in $[0, 1]$. For this purpose, a soft normalization method is used in this paper.

Finally, after the similarity thresholds for all the classes have been learned, the face recognition process can be executed. For a testing face, the partial similarities between it and every training samples are computed under the thresholds of each class. Then, a standard majority voting strategy is used to fuse the obtained intermediate results, and the winner class identifies the testing face.

6 Experiments

6.1 Data and Experimental Setting

Two face databases with large intra-personal variations are used in the experiments, i.e., AR [11] and ORL [3].

The AR face database contains over 4,000 color face images of 126 persons' faces (70 men and 56 women), including frontal view faces with different facial expressions, illuminations, and occlusions (such as 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 2,600 images from 100 different subjects (50 men and 50 women) were used. Some sample images for one subject are shown in Figure 3. Before the recognition process, each image was cropped and resized to 66×48 pixels and then converted to gray-level, which were processed by a histogram equalization algorithm.

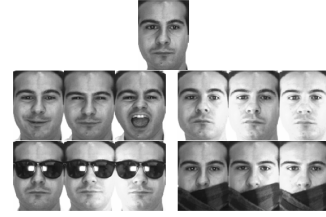


Figure 3. Sample face images taken in the same session for one subject in the AR database.

The ORL face database contains 400 images of 40 persons, and each person has 10 different images. The face images contain significant intra-personal variations caused by rotation, expression and sample size. All images are grayscale and normalized to 112×92 pixels, and the gray-level values of all images are rescaled to $[0, 1]$. Some sample images are shown in Figure 4.



Figure 4. Sample face images for one subject in the ORL database.

The default experimental setting is:

(1) For AR database, the first 7 images per person taken in the first session were used for training, and the remaining 19 images per person were used for testing. In other words, we used 700 training samples and 1900 testing samples in total. The default image partition size is 3×4 pixels and the SOM map size is 94×23 .

(2) For ORL database, the first 5 images per person were used for training and the remaining 5 images per person for testing. The default sub-block size is 4×4 and the SOM map size is 106×17 .

(3) For both database, the threshold search granularity h is set to 100.

6.2 Comparison with Other Distance Measures

One goal of our experiments was to assess the relative performance of the proposed partial distance as a distance measure in face recognition. To this extent we evaluated the nearest neighbor classification using the proposed partial distance, and compared its performance to the nearest neighbor classification using other distance measures.

In particular, four PCA-based non-metric distance measures, the Image Euclidean Distance (IMED), the Relevant Component Analysis distance (RCA), and the traditional Euclidean distance (EU) were chosen for comparison. For RCA distance, since the training set is fully labelled, its chunklets correspond uniquely and fully to the classes. The four PCA-based non-metric distance measures include simplified Mahalanobis (SM), weighted angle-based distance (WA), modified squared Euclidean distance (SE) and angle-based distance between whitened feature vectors (AW). Detailed description of these four distance measures can be found in [13]. According to [13], these four distance measures achieved the best recognition results among 14 distance measures compared in the context of PCA transformation.

The results are shown in Figure 5. Figure 5 reveals that the proposed partial distance measure (PD) significantly outperforms all the compared distance measures consistently on both databases. In particular, the top 1 recognition rates of PD on ORL and AR are 97.0% and 74.6%, respectively, while the best results yielded by the compared distances are 91.5% and 45.8%, respectively. These observations suggest that the proposed partial distance can help filter those over-deformed local facial regions, thus making the robustness against large intra-personal variations be improved.

6.3 Comparison with State-of-the-Art Face Recognition Techniques

In another series of experiments, we compared several state-of-the-art face recognition techniques with the PD-based method. The compared algorithms include the Bayesian method, two kernel-based methods (KPCA and KFLDA), and three popular subspace/manifold algorithms (Eigenface, Fisherface and Lapalacianface). The benchmark algorithm was the direct template matching method

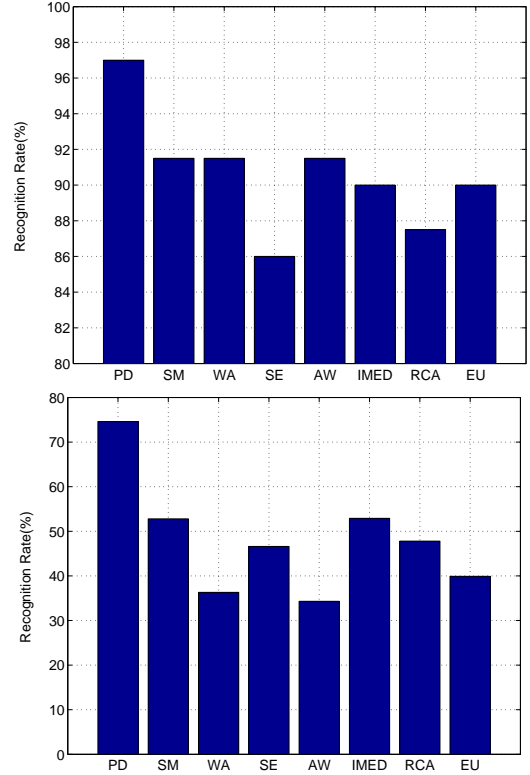


Figure 5. Comparing PD with other 7 distance measures on ORL (top) and AR (bottom).

(1NN with L_2 -norm). In the experiments, 98% information in the sense of reconstruction was kept in the PCA subspaces for all the compared methods. For LDA-like methods (KFLDA, Fisherface and Lapalacianfaces), $(C-1)$ projectors were extracted, where C was the number of total classes. For the two kernel-based methods, Gaussian and polynomial kernels with fine-tuned parameters were tried and the best results were presented.




Table 1 shows the comparison results. Clearly the PD-based method significantly outperforms the direct template matching method and other compared algorithms. These results indicate that although those most current algorithms compared here could yield good performance, their performance still critically depends on the proper estimation of the similarity relationship between face images in the input space. When the intra-personal variation in a database becomes complex, the reliability and accuracy of those algorithms may be seriously affected. The experimental results reported here also reveal that using local patches without selection could not yield good performance, which verifies that the the superior performance of the proposed method is due to the use of the non-metric similarity measure.

For further validation, we detailed the experimental re-

Table 1. Comparing the recognition rates (%) of the PD-based method with state-of-the-art face recognition techniques.

Dataset	PD	Bayesian	Kernel-Methods		Subspace/Manifold Methods			1NN
			KPCA	KFLDA	Eigenface	Fisherface	Lapalacianface	
ORL	97.0	82.0	87.0	84.5	88.5	81.5	83.0	90.0
AR	74.6	52.8	36.3	46.6	34.3	52.9	47.8	39.9

Table 2. Comparing the recognition rates (%) of the PD-based method with state-of-the-art face recognition techniques on face images with large variations. (S1: images taken at the first session, S2: images taken at the second session.)

Face Images	PD	Bayesian	Kernel Methods		Other Subspace/Manifold Methods			1NN
			KPCA	KFLDA	Eigenface	Fisherface	Lapalacianface	
S2 	88.3	83.8	75.3	86.3	74.3	84.8	84.5	79.3
	86.7	85.0	70.3	87.0	70.7	84.7	85.7	76.7
S1 	98.3	42.3	41.3	52.0	34.3	39.0	47.0	48.3
	79.0	21.7	9.0	21.0	8.0	60.7	32.3	10.3
S2 	64.0	72.0	16.3	25.0	15.0	18.7	23.7	19.0
	26.7	12.0	5.7	8.0	3.7	29.3	13.7	6.0

sults on AR in Table 2 by showing the robustness against expression, illumination and occlusions. Table 2 clearly shows the strength of the proposed method.

6.4 Comparison with Other Part-Based Techniques

We also compared the PD-based method with other part-based face recognition techniques, such as NMF [8] and Martinez’s method [9, 10], on AR. NMF learns localized features that can be added together to reconstruct the whole image. However, the learned parts are still represented as high-dimensional vectors in the input space. Martinez explicitly modelled the subspace for each local parts with a parametric model. He found that the recognition accuracy can be improved by incorporating the motion estimation process in the model [10]. In this series of experiments, we used the neutral faces as prototypes, and faces with different expressions (i.e., happy, angry and screaming) for testing. The recognition rates for images of each kind of these facial expressions are shown in Figure 6. It can be found that the

proposed method performs better than both the compared methods consistently on all the expressions.

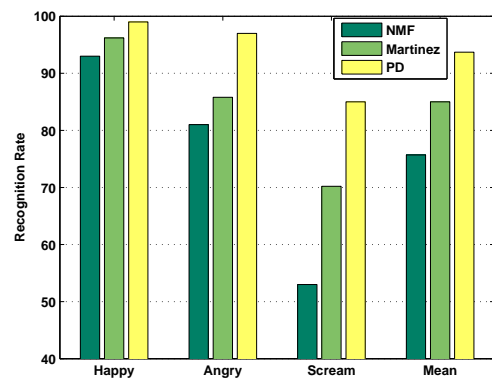


Figure 6. Comparing the recognition rate (%) of the PD-based method with NMF and Martinez’s method on expression variations.

6.5 Effect of Different Sub-block Sizes

Here we take a brief discussion on the influence of sub-block size on the recognition performance. The choice of the sub-block size reflects a balance between generalization and specialization. Generally, as the sub-block gets smaller, the degree of generalization grows higher but the degree of specialization becomes lower. So, the block size should be neither very big nor very small.

In order to verify this statement, we partitioned the original ORL face images with different sizes from smaller to larger and then ran the PD-based method on them. The sub-block sizes are 4×4 , 8×4 , 16×4 , 4×23 and their corresponding recognition performances are 97.0%, 96.5%, 92.5%, 92.0, respectively. These results indicate that a relatively smaller size is preferred in practice than a relatively larger one. However, a too small block size may also be harmful due to the noise it may introduce. Considering the image size (112×92), a block size of 4×4 is not very small but 1×1 is really too small to get a good performance.

7 Conclusion

The main contribution of the paper is the proposal of learning a non-metric partial similarity measure with the maximal margin criterion for robust face recognition. This method simulates the basic visual information processing mechanism of human beings, and potentially allows us understand better the hidden semantic similarity among intra-personal face images. Moreover, to make the method practical, we propose to use the SOM-based embedding technique to deal with the computational problem. Future work includes studying the general applicability of the proposed method, beyond the field of face recognition.

Acknowledgment

We want to thank the anonymous reviewers for their helpful comments and suggestions. We also want to thank Dr. Aleix M. Martinez for providing the AR database. This work was supported by the National Natural Science Foundation of China under Grant No. 60271017, the National Science Fund for Distinguished Young Scholars of China under Grant No. 60325207, and the Jiangsu Science Foundation under Grant No. BK2005122.

References

- [1] B. H. Aharon, T. Hertz, N. Sental, and D. Weinshall. Learning a Mahalanobis metric with side information. *Journal of Machine Learning Research*, 6:937–965, 2005.
- [2] M. S. Bartlett. *Face Image Analysis by Unsupervised Learning*. Kluwer, Boston, 2001.
- [3] P. Belhumeur, J. Hespanha, and D. Kriegman. Eigenfaces vs. Fisherfaces: Recognition using class specific linear projection. *IEEE Trans. Pattern Analysis and Machine Intelligence*, 19(7):711–720, 1997.
- [4] X. F. He, X. C. Yan, Y. Hu, P. Niyogi, and H. J. Zhang. Face recognition using Laplacianfaces. *IEEE Trans. Pattern Analysis and Machine Intelligence*, 27(3):328–340, 2005.
- [5] G. R. Hjaltason and H. Samet. Properties of embedding methods for similarity searching in metric spaces. *IEEE Trans. Pattern Analysis and Machine Intelligence*, 25(5):530–549, 2003.
- [6] D. W. Jacobs, D. Weinshall, and Y. Gdalyahu. Classification with non-metric distances: Image retrieval and class representation. *IEEE Trans. Pattern Analysis and Machine Intelligence*, 22(6):583–600, 2000.
- [7] T. Kohonen. *Self-Organizing Map*. Springer, Berlin, 2nd edition, 1997.
- [8] D. D. Lee and H. S. Seung. Learning the parts of objects by non-negative matrix factorization. *Nature*, 401:788–791, 1999.
- [9] M. Martinez. Recognizing imprecisely localized, partially occluded, and expression variant faces from a single sample per class. *IEEE Trans. Pattern Analysis and Machine Intelligence*, 24(6):748–763, 2002.
- [10] M. Martinez. Matching expression variant faces. *Vision Research*, 43(9):1047–1060, 2003.
- [11] M. Martinez and R. Benavente. The AR face database. Technical Report 24, CVC, 1998.
- [12] B. Moghaddam and A. Pentland. Probabilistic visual learning for object representation. *IEEE Trans. Pattern Analysis and Machine Intelligence*, 19(7):696–710, 1997.
- [13] V. Perlibakas. Distance measures for PCA-based face recognition. *Pattern Recognition Letters*, 25(6):711–724, 2004.
- [14] S. Santini and R. Jain. Similarity measures. *IEEE Trans. Pattern Analysis and Machine Intelligence*, 21(9):871–883, 1999.
- [15] 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 Trans. Neural Networks*, 16(4):875–886, 2005.
- [16] A. Tversky. Features of similarity. *Psychological Review*, 84(4):327–352, 1977.
- [17] P. Viola and M. Jones. Robust real-time object detection. *International Journal of Computer Vision*, 57(2):137–154, 2004.
- [18] J. Wang, Y. Zhang, and J. Feng. On the Euclidean distance of images. *IEEE Trans. Pattern Analysis and Machine Intelligence*, 27(8):1334–1339, 2005.
- [19] E. P. Xing, A. Y. Ng, M. I. Jordan, and S. Russell. Distance metric learning, with application to clustering with side-information. In S. Becker, S. Thrun, and K. Obermayer, editors, *Advances in Neural Information Processing Systems 15*, pages 505–512. MIT Press, Cambridge, MA, 2003.
- [20] M. H. Yang. Kernel eigenfaces vs. kernel fisherfaces: Face recognition using kernel methods. In *Proc. the 5th IEEE International Conference on Automatic Face and Gesture Recognition*, pages 215–220, Washington, DC, 2002.
- [21] W. Zhao, R. Chellappa, P. J. Phillips, and A. Rosenfeld. Face recognition: A literature survey. *ACM Computing Survey*, 34(4):399–485, 2003.