

Face Recognition Using Face Patch Networks

Chaochao Lu

Deli Zhao

Xiaou Tang*

Department of Information Engineering, The Chinese University of Hong Kong

{cclu, dlzhao, xtang}@ie.cuhk.edu.hk

Abstract

When face images are taken in the wild, the large variations in facial pose, illumination, and expression make face recognition challenging. The most fundamental problem for face recognition is to measure the similarity between faces. The traditional measurements such as various mathematical norms, Hausdorff distance, and approximate geodesic distance cannot accurately capture the structural information between faces in such complex circumstances. To address this issue, we develop a novel face patch network, based on which we define a new similarity measure called the random path (RP) measure. The RP measure is derived from the collective similarity of paths by performing random walks in the network. It can globally characterize the contextual and curved structures of the face space. To apply the RP measure, we construct two kinds of networks: the in-face network and the out-face network. The in-face network is drawn from any two face images and captures the local structural information. The out-face network is constructed from all the training face patches, thereby modeling the global structures of face space. The two face networks are structurally complementary and can be combined together to improve the recognition performance. Experiments on the Multi-PIE and LFW benchmarks show that the RP measure outperforms most of the state-of-art algorithms for face recognition.

1. Introduction

Over the past two decades, face recognition has been studied extensively [10, 14, 19, 33, 4, 6, 34, 3, 30, 16]. However, large intra-personal variations, such as pose [24, 35], illumination [24, 8], and expression [2, 24], remain challenging for robust face recognition in real-life photos. In Figure 1, for example, A and A' are two images of the same person with different poses and illuminations. A and

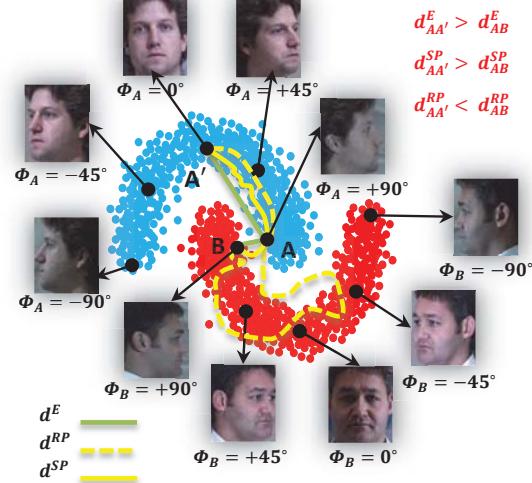


Figure 1. Illustration of the superiority of our random path (RP) measure over other measures (for example, Euclidean (E) measure and the shortest path (SP) measure). Due to the large intra-personal variations (e.g., pose, illumination, and expression), there may be underlying structures in face space (denoted by the red and blue clusters). For three face images A, B, and A' of two different persons, the distances are $d_{AA'}^E > d_{AB}^E$ and $d_{AA'}^{SP} > d_{AB}^{SP}$ if measured by Euclidean measure (solid green line) and the shortest path measure (solid yellow line). In other words, A is more similar to B than to A'. Incorrect decisions are usually made because the intra-personal variation is much larger than the inter-personal variation. If we consider their underlying structures and compute their similarity by our random path measure (dashed yellow line), we get $d_{AA'}^{RP} < d_{AB}^{RP}$. The correct decision can be made. Note that this figure is only for the purpose of schematic illustration. In the real experiment, we use facial patches instead of the whole face.

B are from two different persons with the same pose and illumination. The appearances of A and B are more similar to each other than A is to A', which may confuse most existing face recognition algorithms.

Classical measurement approaches for face recognition have several limitations, which have restricted their wider applications in the scenarios of large intra-personal variations. Seminal studies in [27, 23, 20, 22] have revealed

*This work is supported by the General Research Fund sponsored by the Research Grants Council of the Hong Kong SAR (Project No. CUHK 416312 and CUHK 416510) and Guangdong Innovative Research Team Program (No.201001D0104648280).

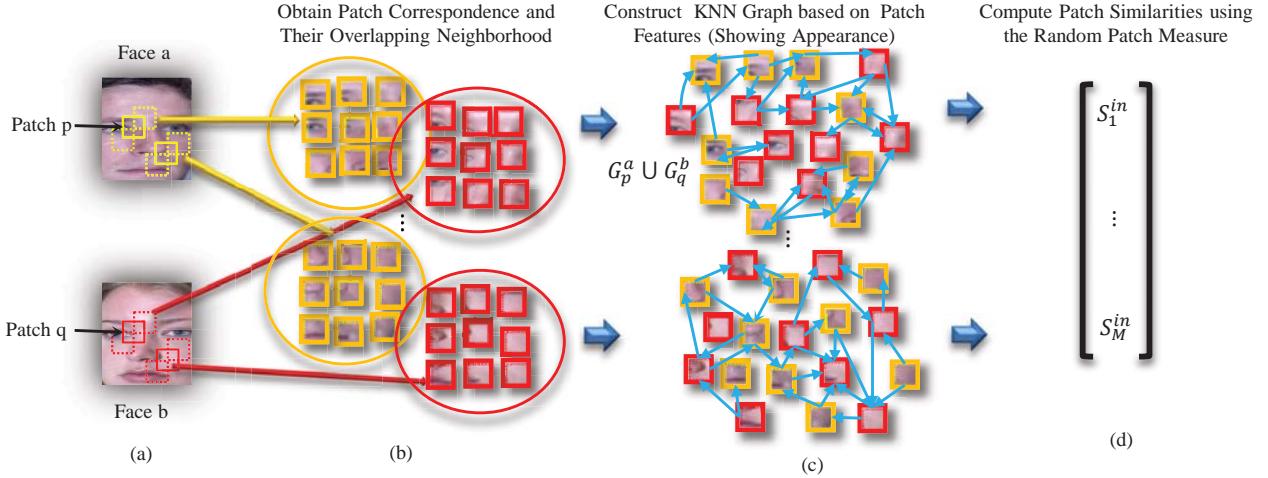


Figure 2. The proposed in-face network pipeline.

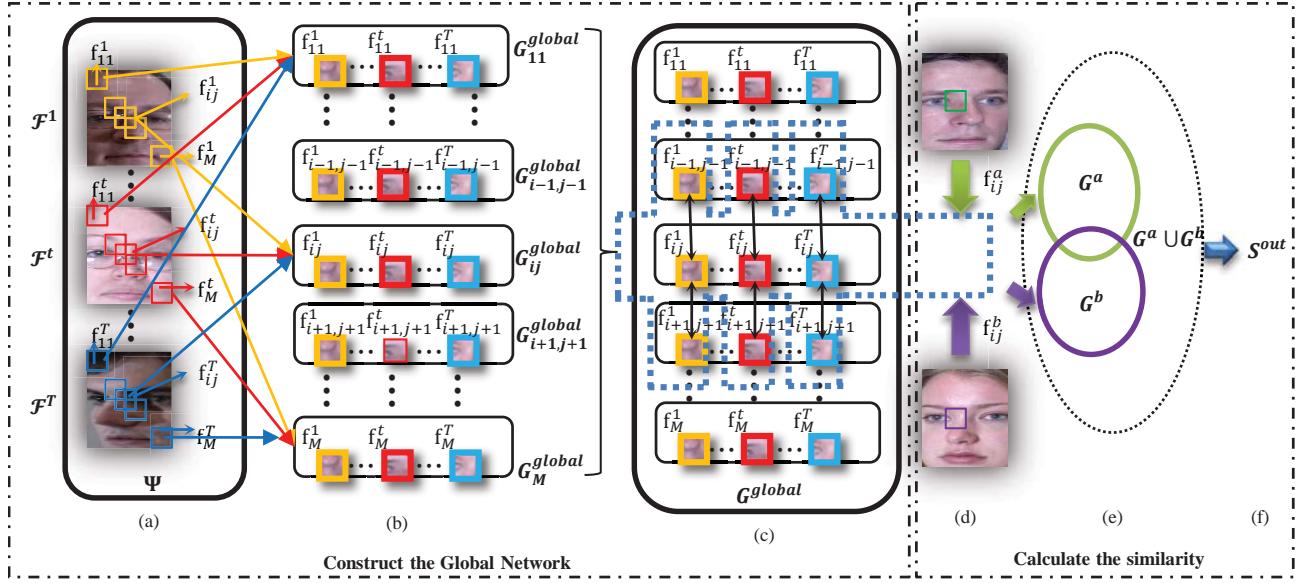


Figure 3. The proposed out-face network pipeline.

that the diverse distributions of face images for one person may form the underlying manifold structures. In fact, these distributions generally have different densities, sizes, and shapes, due to the high complexity in face data. In addition, noise and outliers are often contained in face space. The current measurement approaches fail to tackle all of these challenges. Many methods based on (dis-)similarity measures [1, 25, 32, 10] directly use pairwise distances to compute the (dis-)similarities between faces, which cannot capture the structural information for the high-quality discrimination, as shown in Figure 1. Although some studies [29, 27, 21] apply the structural information in measurements, the developed algorithms are, generally speaking, sensitive to noise and outliers. For instance,

computing the length of the shortest path in a network is very sensitive to noisy nodes.

This paper reports on a new face similarity measure called random path (RP) measure, which was designed to overcome the above-mentioned problems. We first construct two novel face patch networks: the in-face network and the out-face network, as shown in Figures 2 and 3. In our study, faces are divided into multiple overlapping patches of the same size. The in-face network is defined for any pair of faces. For each pair of faces, at each patch location, we use the two corresponding patch pairs and their eight neighboring patches to form a KNN graph, which we call the in-face network. For each such in-face network, we propose a random path (RP) measure as the patch similarity

of the corresponding patch pair. Given a network, all paths between arbitrary two nodes are integrated by a generating function. The RP measure includes all paths of different lengths in the network, which enables it to capture more discriminative information in faces and significantly reduce the effect of noise and outliers. For a pair of faces with M patches, therefore, we can compute M RP measures to form the similarity feature vector between the two faces. Since the network is only constructed within two faces in this approach, we call it the in-face network.

The out-face network is built in a similar fashion. Instead of using local neighboring patches to form the network, for each patch we search a database of face patches and find similar patches in the same location neighbors of the patch to form the patch pair network. Since the search is conducted globally over the training space, the out-face network captures more global structural information. Because the two networks describe the local and global structural information respectively, the similarities derived from the RP measure on these two networks can be combined to boost the recognition performance. By means of the RP measure on the in-face and outface networks, our RP measure performs significantly better than existing measures for face verification on two challenging face datasets, LFW [11] and Multi-PIE [9]. In addition, our method outperforms most of the current appearance-based methods [15, 4, 18, 7].

2. Related Work

From the point of view of data structures, there are two types of measures for face recognition: non-structure-based measures [1, 25, 32, 10, 4] and structure-based measures [29, 27, 21]. Here, the structure means that data points may lie in some underlying manifolds and the distance between data points cannot be accurately measured by a straight line or its analogous variants. Since the face space is often structured - that is, nonlinear [27, 23, 20, 22] - the distance between faces cannot be precisely measured by non-structured measures such as Euclidian distance, Manhattan distance, Pearson's coefficient of correlation, Hausdorff distance, and Chi-square distance, or by linear subspaces derived from linear projection methods such as Principal Component Analysis (PCA) [28] and Linear Discriminant Analysis (LDA) [2, 31, 30].

Studies on similarity measure in face recognition mainly focus on the non-structured metrics. In [1], for example, a Manhattan-like measure was proposed, which is a weighted sum of Chi-square distances. In [4], the component similarity is measured by L_2 distance between the corresponding descriptors of the face pair. A partial Hausdorff distance measure was defined in [25], where a pixel in one face could be matched with any other pixel with the same local binary pattern in the other face. Similarly, in [10], a robust

elastic and partial matching metric was presented, where each descriptor in one face is matched with its spatial neighboring descriptors in another face and the minimal distance is regarded as their dis-similarity. Although some measures, such as the one in [10], have taken spatial neighborhood information into consideration, they all directly use pairwise distances to measure similarities without structural information.

For the structural information based measures, one of the representative works is the Isomap algorithm in manifold learning [27]. Isomap models the structural proximity between two data points by the geodesic distance that can be approximated by the shortest path length. Based on geodesic distances, spectral low-dimensional embeddings of data correctly unfold the underlying manifold. The idea was applied in [29] for face recognition by proposing the Manifold-Manifold Distance (MMD). Instead of measuring two data points, the MMD measures the similarity of two manifolds by finding the shortest distance between component subspaces from the two manifolds. Although these approaches can capture the manifold structures, their performance often degrades significantly with the existence of noise or outliers, because the shortest path is sensitive to noisy perturbations.

3. Modeling Faces with Patch Networks

A face is holistically structured. Even for a patch p cropped from the face, its micro-structure is continuously connected with that of patches around patch p . For instance, the structure of the patch of the eye corner is bridged with that of its neighboring patches, as Figures 2 (a) and (b) show. Therefore, it is more convincing to take the neighboring patches into consideration when locally comparing the similarity between two faces from patch pair. To do this, we sample the r neighboring patches around patch p . Figure 2 (b) shows the patch examples in the case of $r = 8$ for face a and face b . The visual semantics of local patches cropped at the same point for different faces may be highly correlated. For example, the eye corners for different faces are very similar. Therefore, the $2 \times (r + 1)$ neighbor patches may be spatially mixed if we present them as spatial vector points. To model such mixing structures, we resort to complex network. The mixing patches are locally connected according to their spatial positions in feature space by a KNN network. Figure 2 (c) shows a 2-NN network of facial patches. The network constructed between two faces is called the in-face network.

Figure 2 (c) shows that some patches of the same person are not directly linked. To model the structure in such case, we apply paths to connect the distant nodes in the network. Path-based measurements are widely employed in social networks to compute the nodal similarities. For instance, each entry of $(\mathbf{I} - z\mathbf{P})^{-1}$ presents the global

similarity for the corresponding nodes [13], where \mathbf{P} is the adjacency matrix of the network. The nodal similarity is defined by all the connected paths between nodes. With the similarity matrix $(\mathbf{I} - z\mathbf{P})^{-1}$, we can measure the structural compactness of the network by the average similarity between all nodes. To this end, we define the path centrality $C_G = \frac{1}{N} \mathbf{1}^T (\mathbf{I} - z\mathbf{P})^{-1} \mathbf{1}$. The more compact the network is, the larger the path centrality is. With path centrality C_G , it will be easy to measure the similarity of patch pair in the network framework. To make the analysis clearer, we let $G_p^a \cup G_q^b$ denote the network constructed from patch p in face a and patch q in face b , as shown in Figure 2 (a), where G_p^a is the sub-network of patch p and G_q^b is the sub-network of patch q . It is straightforward to know that the more similar the patch p and the patch q are, the more mutually connected paths there are between G_p^a and G_q^b . Therefore, the increment of the path centrality $C_{G_p^a \cup G_q^b}$ will be large over $C_{G_p^a}$ and $C_{G_q^b}$ of the sub-networks, which motivates us to define the random path measure $C_{G_p^a \cup G_q^b} - C_{G_p^a} - C_{G_q^b}$ for measuring the similarity of patch pair. We will present the formulation in Section 3.1.

For many faces, there also exists the correlation between the same components of different faces. For example, Peter's eyes look like Tom's. To model such structural correlation, we further construct a network from the patches at the same positions of all faces. Thus we can build the M networks if there are M different patches in one face, as Figures 3 (a) and (b) illustrate. To integrate these component networks, for each patch in a face, we link it with its r most similar patches in the neighbor position in the training data, as Figures 3 (c) shows. Thus we derive a global network that cover the structural correlation between faces and within faces. We call it the out-face network.

The random path measure will be adopted both in the in-face and out-face networks. So we first present the formulation of the random path measure and then the construction of the in-face and out-face networks follows.

3.1. The Random Path Measure

Let G denote a network with N nodes $\{x_1, \dots, x_N\}$, and \mathbf{P} denote its weighted adjacency matrix. Each entry in the matrix \mathbf{P} is the similarity between associated nodes. For generality, G is assumed to be directed, which means that \mathbf{P} may be asymmetric. A path of length t defined on \mathbf{P} is denoted by $p_t = \{v_0 \rightarrow v_1 \rightarrow \dots \rightarrow v_{t-1} \rightarrow v_t\}$. S_t is the set of all paths of length t . Let T denote the transpose operator of a matrix, $\mathbf{1}$ the all-one vector, and \mathbf{I} the identity matrix.

Inspired by concepts in social network analysis [17], we introduce the definition of path centrality C_G for the network G .

Definition 1 Path Centrality $C_G = \frac{1}{N} \mathbf{1}^T (\mathbf{I} - z\mathbf{P})^{-1} \mathbf{1}$, where $z < 1/\rho(\mathbf{P})$ and $\rho(\mathbf{P})$ is the spectral radius of \mathbf{P} .

The (i, j) entry of the matrix $(\mathbf{I} - z\mathbf{P})^{-1}$ represents a kind of global similarity between node x_i and node x_j . It was first introduced by Katz [13] to measure the degree of influence of an actor in a social network. To make it clear, we expand $(\mathbf{I} - z\mathbf{P})^{-1}$ and view it as a generating matrix function

$$(\mathbf{I} - z\mathbf{P})^{-1} = \mathbf{I} + z\mathbf{P} + z^2\mathbf{P}^2 + \dots = \sum_{t=0}^{\infty} z^t \mathbf{P}^t. \quad (1)$$

Each entry in the matrix \mathbf{P}^t can be written as

$$\mathbf{P}_{i,j}^t = \sum_{p_t \in S_t} \prod_{k=0}^{t-1} \mathbf{P}_{v_k, v_{k+1}}, \quad (2)$$

which is the sum of the products of the weights over all paths of length t that start at node x_i and end at node x_j in G . In machine learning, the global similarity defined by Eq. (2) is also called the semantic similarity [12]. In our framework, the weighted adjacency matrix \mathbf{P} satisfies that each entry is non-negative and each row sum is normalized to 1. Therefore, we can view the entry $\mathbf{P}_{i,j}^t$ as the probability that a random walker starts from node x_i and arrives at node x_j after t steps. From this point of view, the path centrality is to measure the structural compactness of the network G by all paths of all lengths between all the connected nodes in G . Due to the randomness of walks in G , we refer to our measurement as the *random path measure*.

With the definition of path centrality, the RP measure can be naturally used to compute the similarity between two networks. From the definition of path centrality, it makes sense that the two sub-networks in G have the most similar structures in the sense of path centrality if they share the most paths. In other words, from the viewpoint of structural recognition, the two networks are most relevant. Therefore, for two given networks G_i and G_j , the definition of our RP measure can be defined as follows.

Definition 2 Random Path Measure $\Phi_{G_i \cup G_j} = C_{G_i \cup G_j} - (C_{G_i} + C_{G_j})$, is regarded as the similarity between two networks G_i and G_j .

In the definition above, the union path centrality $C_{G_i \cup G_j}$ is written as

$$C_{G_i \cup G_j} = \frac{1}{|G_i \cup G_j|} \mathbf{1}^T (\mathbf{I} - z\mathbf{P}_{G_i \cup G_j})^{-1} \mathbf{1}. \quad (3)$$

where $\mathbf{P}_{G_i \cup G_j}$ is the union adjacency matrix corresponding to the nodes in G_i and G_j . The RP measure $\Phi_{G_i \cup G_j}$ embodies the structural information about all paths between G_i and G_j . In order to understand the definition intuitively, we consider a case shown in Figure 4. C_{G_i} and C_{G_j} measure the structural information in G_i and G_j , respectively.

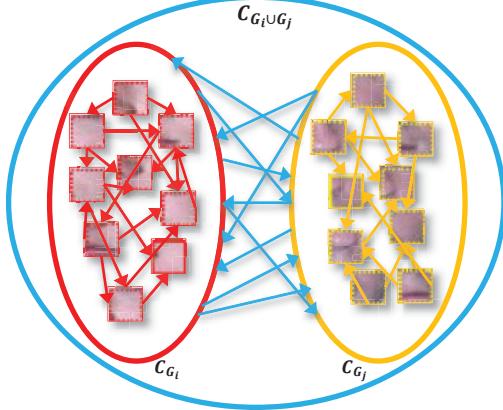


Figure 4. Illustration of the random path measure. Paths in G_i and G_j are denoted by red and yellow arrows, respectively. Paths between G_i and G_j are denoted by blue arrows.

$C_{G_i \cup G_j}$ measures not only the structural information within G_i and G_j , but also that through all paths between G_i and G_j . The larger the value of $\Phi_{G_i \cup G_j}$, the more structural information the two networks share, meaning that these two networks have more similar structures. Therefore, $\Phi_{G_i \cup G_j}$ can be exploited to measure the structural similarity between two networks.

The RP measure takes all paths between two networks into consideration to measure their similarity, not only the shortest path such as [29, 27]. Therefore, our measure is robust to noise and outliers. Besides, we take the average value of nodal centrality $(\mathbf{I} - z\mathbf{P})^{-1}\mathbf{1}$. With this operation, the structural information of network is distributed to each node, which means that the RP measure is also insensitive to multiple distributions and multiple scales.

3.2. In-face Network

Figure 2 presents the in-face network pipeline. We densely partition the face image into $M = K \times K$ overlapping patches of size $n \times n$ ($n = 16$ in our settings). We set 8-pixel overlap in this paper. We apply a local image descriptor to extract features for each patch of size $n \times n$. Therefore, each face is represented by a set of $M = K \times K$ feature vectors formed from a local image descriptors

$$\mathcal{F} = \{\mathbf{f}_{11}, \dots, \mathbf{f}_{ij}, \dots, \mathbf{f}_M\}, \quad (4)$$

where \mathbf{f}_{ij} is the feature vector of the patch located at (i, j) in the face. \mathcal{F}^a and \mathcal{F}^b denote the feature sets of face a and face b , respectively. To build an in-face network for the patches at (i, j) , we take \mathbf{f}_{ij}^a in \mathcal{F}^a and \mathbf{f}_{ij}^b in \mathcal{F}^b . At the same time, the r neighbors of \mathbf{f}_{ij}^a around (i, j) are also taken. The same operation is also performed for \mathbf{f}_{ij}^b . We set $r = 8$ in this paper. Thus, we get the $(2 + 2r)$ feature vectors of patches that are utilized to construct a KNN network G_{ij} for the patch pair of \mathbf{f}_{ij}^a and \mathbf{f}_{ij}^b . Its weighted adjacency matrix

is denoted by $\mathbf{P}_{G_{ij}}$. Therefore, the adjacency matrix $\mathbf{P}_{G_{ij}^a}$ of the network G_{ij}^a corresponding to \mathbf{f}_{ij}^a and its r neighbors is the sub-matrix of G_{ij} identified by the indices of \mathbf{f}_{ij}^a and its r neighbors. Similarly, we can get G_{ij}^b and its adjacency matrix $\mathbf{P}_{G_{ij}^b}$. For better understanding, we define $G_{ij} = G_{ij}^a \cup G_{ij}^b$, which means the set of nodes of G_{ij} are the union of that of G_{ij}^a and G_{ij}^b . For the patch pair of \mathbf{f}_{ij}^a and \mathbf{f}_{ij}^b , we calculate their path centralities as follows:

$$\begin{aligned} C_{G_{ij}^a} &= \frac{1}{|G_{ij}^a|} \mathbf{1}^T (\mathbf{I} - z\mathbf{P}_{G_{ij}^a})^{-1} \mathbf{1}, \\ C_{G_{ij}^b} &= \frac{1}{|G_{ij}^b|} \mathbf{1}^T (\mathbf{I} - z\mathbf{P}_{G_{ij}^b})^{-1} \mathbf{1}, \\ C_{G_{ij}^a \cup G_{ij}^b} &= C_{G_{ij}} = \frac{1}{|G_{ij}|} \mathbf{1}^T (\mathbf{I} - z\mathbf{P}_{G_{ij}})^{-1} \mathbf{1}. \end{aligned} \quad (5)$$

Applying the RP measure gives the similarity measure of the patch pair

$$S_{ij}^{in} = \Phi_{G_{ij}^a \cup G_{ij}^b} = C_{G_{ij}^a \cup G_{ij}^b} - (C_{G_{ij}^a} + C_{G_{ij}^b}). \quad (6)$$

Analogous to this manipulation, the similarities of M patch pairs from \mathcal{F}^a and \mathcal{F}^b can be derived. Padding them as a similarity vector

$$\mathbf{s}^{in} = [S_{11}^{in}, \dots, S_{ij}^{in}, \dots, S_M^{in}]^T. \quad (7)$$

completes the process of applying the RP measure on the in-face network for two face images.

We refer to the network presented above as the in-face network because the network is only constructed within two face images. Only the structural information of patch pair and their neighborhoods is considered; therefore, the in-face network mainly conveys the local information.

3.3. Out-face Network

The proposed out-face network pipeline is shown in Figure 3. Unlike the in-face network, the construction of the out-face network requires the training data in an unsupervised way. The patch division and feature extraction is performed in the same way as in Section 3.2. Suppose that T is the number of face images in the training set. Write the feature set as

$$\Psi = \{\mathcal{F}^1, \dots, \mathcal{F}^T\}, \quad (8)$$

where \mathcal{F}^t is the feature set of the t -th face. We first adopt all the feature vectors $\{\mathbf{f}_{ij}^1, \dots, \mathbf{f}_{ij}^T\}$ at (i, j) in the training set to construct a KNN network G_{ij}^{global} . In this way, we can construct M independent G_{ij}^{global} , meaning that there is no connection between them. Further, to preserve the structural proximity between \mathbf{f}_{ij}^t and its neighbors at (i, j) in each face, we connect \mathbf{f}_{ij}^t with all of its 8 neighbors.

Here by ‘‘connect’’ we mean when a patch is selected, all its r neighbors will also be selected. Therefore, by the connections of neighborhoods, the M independent G_{ij}^{global} are linked together to form the final global network G^{global} with the weighted adjacency matrix \mathbf{P}^{global} .

Given a test face image a , we search its r^{NN} most similar patches in G_{ij}^{global} for each \mathbf{f}_{ij}^a , and then for each selected patch, we also select its 8 neighbor patches together to form the initial G^a . This search method can guarantee that the acquired similar patches are among the spatially semantic neighbors of \mathbf{f}_{ij}^a in other face images. Thus, $(r^{NN} + 1) \times M$ patch nodes are finally selected from G^{global} . We delete some duplicates from them and use the remaining nodes to extract the sub-network G^a from G^{global} with its corresponding sub-matrix \mathbf{P}_{G^a} from \mathbf{P}^{global} . G^b and \mathbf{P}_{G^b} can be acquired in the same way for face b . By merging nodes in G^a and G^b , we can draw the union network $G^a \cup G^b$ and its adjacency matrix $\mathbf{P}_{G^a \cup G^b}$ from G^{global} and \mathbf{P}^{global} .

After acquiring \mathbf{P}_{G^a} , \mathbf{P}_{G^b} , and $\mathbf{P}_{G^a \cup G^b}$ for face a and face b , it is straightforward to compute their path centralities: C_{G^a} , C_{G^b} , and $C_{G^a \cup G^b}$ according to Definition 1. We then utilize the RP measure to calculate their similarity

$$s^{out} = \Phi_{G^a \cup G^b}. \quad (9)$$

s^{out} describes the structural information of two face images from the global view.

Since the construction of this network requires the training data and the each test face needs to be projected on it, we call the network the out-face network. Searching for the nearest neighbors for each patch is fast because the search operation is only made in G_{ij}^{global} instead of G^{global} .

3.4. The Fusion Method

From the analysis above, it is clear that the in-face network and the out-face network are structurally complementary. To improve the discriminative capability of the networks, we present a simple fusion method to combine them

$$\mathbf{s}^{final} = [\alpha \mathbf{s}^{in}, (1 - \alpha) \mathbf{s}^{out}], \quad (10)$$

where \mathbf{s}^{final} is the combined similarity vector of two face images, and α is a free parameter learned from the training data. This fusion method can effectively combine the advantages of the in-face network and the out-face network. We feed \mathbf{s}^{final} to the linear SVM [5] to train a classifier for recognition.

3.5. Weighted Adjacency Matrix

The weight $\mathbf{P}(i, j)$ of the edge connecting node x_i and node x_j in the network is defined as

$$\mathbf{P}(i, j) = \begin{cases} \exp\left(-\frac{dist(x_i, x_j)^2}{\sigma^2}\right), & \text{if } x_j \in \mathcal{N}_i^K \\ 0, & \text{otherwise} \end{cases} \quad (11)$$

where $dist(x_i, x_j)$ is the pairwise distance between x_i and x_j , \mathcal{N}_i^K is the set of KNNs of x_i , and $\sigma^2 = \frac{1}{nK} [\sum_{i=1}^n \sum_{x_j \in \mathcal{N}_i^K} dist(x_i, x_j)^2]$. To get the transition probability matrix, we perform $\mathbf{P}(i, j) \leftarrow \mathbf{P}(i, j) / \sum_{j=1}^n \mathbf{P}(i, j)$.

4. Experiments

In this section, we conduct experiments on face verification to validate the effectiveness of our RP measure based on the in-face and out-face networks. The face data we use are two widely used face databases: the Multi-PIE dataset [9] and the LFW dataset [11]. The Multi-PIE dataset contains face images from 337 subjects under 15 view points and 19 illumination conditions in four recording sessions. Unlike the Multi-PIE dataset, the LFW dataset contains 13,233 uncontrolled face images of 5,749 public figures of different ethnicity, gender, age, etc.

In our settings, according to subject identities, the Multi-PIE dataset is divided into three parts: S_1 (ID 1-100), S_2 (ID 101-300), and S_3 (ID 301-346). We collect a new dataset S_{train} by randomly selecting 3000 face images from S_1 . The S_{train} is applied to construct the global network G^{global} employed in Section 4.1 and 4.2. From S_2 , we randomly select 10 mutually disjoint folders with 500 intra-personal and 500 extra-personal pairs in each folder. This dataset will be used for testing in Section 4.2. The remaining S_3 is applied in Section 4.1. We also randomly select 10 mutual disjoint folders with 100 intra-personal and 100 extra-personal pairs from S_3 to tune the optimal parameters. For the LFW dataset, we follow the restricted protocol of the LFW benchmark for evaluation [11]. To perform the fair comparison with the recent algorithms in face recognition, we follow the procedures in [4] to crop faces and each cropped face is resized to 84×96 pixels with the eyes and mouth corners aligned.

To verify the performance of the proposed RP measure, we compare our algorithms mainly to the widely used measures, including Euclidean distance, Chi-square distance, Hausdorff distance, Hua *et al.*’s method [10], and the shortest path [26, 29], on four popular descriptors in face recognition: LBP [18], HOG [7], Gabor [32], and LE [4].

4.1. Tuning Parameters

Since our approaches involve some free parameters, we first determine the optimal parameters used in our approaches on the randomly selected face image collection from S_3 . Our approaches involve four important parameters¹. The first two parameters are the number K^{in} of nearest

¹There are also three relatively unimportant parameters: the size of the patch (n), the patch sampling step (s), and the number of the patch’s neighbors (r). Intuitively, the size and the step should not be too large or too small; so it is good that $n = 16$ and $s = 8$. Usually, the patch is only relative to its eight surrounding neighbors, so r is set to 8.

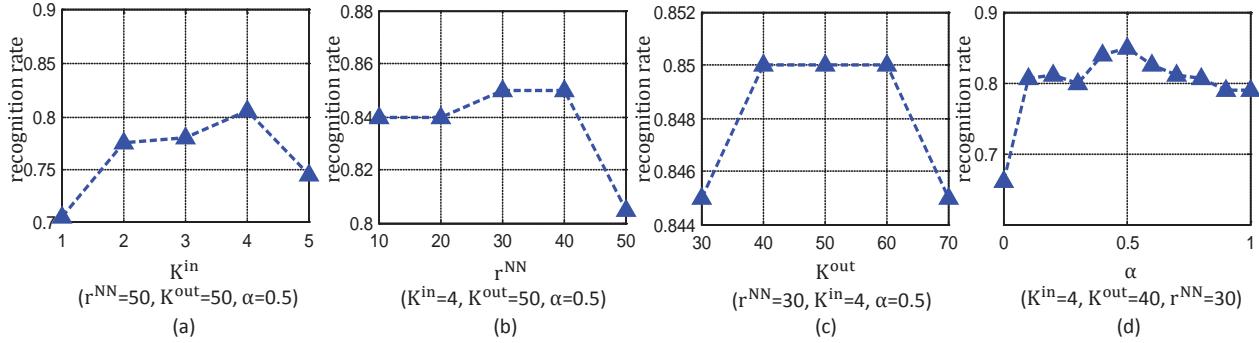


Figure 5. Setting parameters. There are four important parameters in our approaches. We tune one of four parameters while keeping the other parameter unchanged.

Dataset	Recog. rate on Multi-PIE				Recog. rate on LFW			
	Descriptor	LBP	HOG	Gabor	LE	LBP	HOG	Gabor
l_1	70.8 \pm 0.8	72.8 \pm 0.7	72.1 \pm 0.7	80.1 \pm 0.7	66.9 \pm 0.6	66.3 \pm 0.5	61.4 \pm 0.3	73.4 \pm 0.4
Euclidean	66.9 \pm 0.4	76.9 \pm 0.8	72.7 \pm 0.6	75.8 \pm 0.6	62.2 \pm 0.9	68.2 \pm 0.6	49.5 \pm 0.5	65.2 \pm 0.8
Chi-Square	70.1 \pm 0.6	75.9 \pm 0.9	49.9 \pm 0.1	73.6 \pm 0.6	67.2 \pm 0.7	68.4 \pm 0.7	50.1 \pm 0.1	73.4 \pm 0.4
Hausdorff	61.9 \pm 0.7	76.1 \pm 0.9	68.9 \pm 0.7	57.1 \pm 0.8	50.8 \pm 0.5	64.9 \pm 0.7	61.8 \pm 0.4	51.7 \pm 0.3
Hua <i>et al</i> [10]	61.4 \pm 0.3	69.9 \pm 0.4	70.9 \pm 0.3	73.8 \pm 0.7	64.3 \pm 0.8	65.7 \pm 0.7	64.1 \pm 0.9	67.4 \pm 0.9
Shortest Path [26, 29]	51.2 \pm 0.1	50.3 \pm 0.3	52.1 \pm 0.3	50.4 \pm 0.5	50.1 \pm 0.1	50.7 \pm 0.4	52.3 \pm 0.5	50.4 \pm 0.1
Our in-face	81.3 \pm 0.4	85.1 \pm 0.6	83.6 \pm 0.3	88.1 \pm 0.7	77.1 \pm 0.1	74.2 \pm 0.3	71.5 \pm 0.4	84.4 \pm 0.2
Our out-face	71.8 \pm 0.6	77.8 \pm 0.7	74.1 \pm 0.6	81.3 \pm 0.2	67.4 \pm 0.4	69.5 \pm 0.2	65.6 \pm 0.1	79.5 \pm 0.1
Our fusion	85.3 \pm 0.1	88.6 \pm 0.5	87.4 \pm 0.1	92.3 \pm 0.3	80.1 \pm 0.3	76.3 \pm 0.1	74.4 \pm 0.6	86.7 \pm 0.4

Table 1. Results on Multi-PIE and LFW. In the experiment, the out-face networks for Multi-PIE and LFW are the same and constructed from the Multi-PIE faces.

neighbors in the construction of the in-face network and K^{out} in the construction of the out-face network. K^{in} and K^{out} play a very important role in our approaches because they directly determine the structures of the in-face network and the out-face networks. The third parameter is r^{NN} , which is the number of nearest neighbors for each patch of a test face in G^{global} . It controls the inter-personal complexity of the out-face network. The fourth parameter is the weighting parameter α in the fusion method. To balance the importance of similarities yielded by the in-face network and the out-face network, this parameter should be chosen carefully.

In this section, we conduct four experiments to explore the effects of the four parameters. In all of the experiments, we extract LBP features for facial patches. When tuning one of the four parameters, we keep the other three ones unchanged. For example, in Figure 5 (a), we fix that $K^{out} = 50$, $r^{NN} = 50$, and $\alpha = 0.5$. Then, the optimal value of K^{in} is acquired, as $K^{in} = 4$, when the algorithm achieves the best performance. The adjustment of K^{out} , r^{NN} , and α are shown in Figures 5 (b), (c), and (d). In Figure 5 (b), the verification performance coincides when $r^{NN} = 30$ and $r^{NN} = 40$. For fast computation, $r^{NN} = 30$ is chosen. Similarly, we get $K^{out} = 40$ from Figure 5 (c). Therefore, we determine that $K^{in} = 4$, $K^{out} = 40$, $r^{NN} = 30$, and $\alpha = 0.5$ in our parameter settings.

4.2. Results on Multi-PIE and LFW

Table 1 provides the results of our RP measure and other measures for comparison on two face database benchmarks. The results clearly show that our RP measure can dramatically improve the recognition performance of the four descriptors. In addition, the results in the last three rows in Table 1 effectively verify that the in-face network and the out-face network are complementary, because the recognition performance of the combined network can be improved over that of the in-face and out-face networks. Our RP measure is a general similarity measurement and can be applied to improve any appearance-based approach.

To further demonstrate the robust performance of our method, we present the verification results on the LFW dataset with the outside training data. 10,000 face images from the Multi-PIE and Pugfig83 [14] databases are adopted to construct the out-face network. The feature vector for each patch is the combined features of LBP, HOG, Gabor, and LE. As shown in Figure 6, our method performs best when all of the methods are directly performed on the original faces. The best performance for LFW is 93.3% reported in [3]. Their method applies the accurate face alignment and warp with the human-labeled locations of 95 face parts for 20,639 face images. If all images in LFW are aligned with global affine transformations based on the

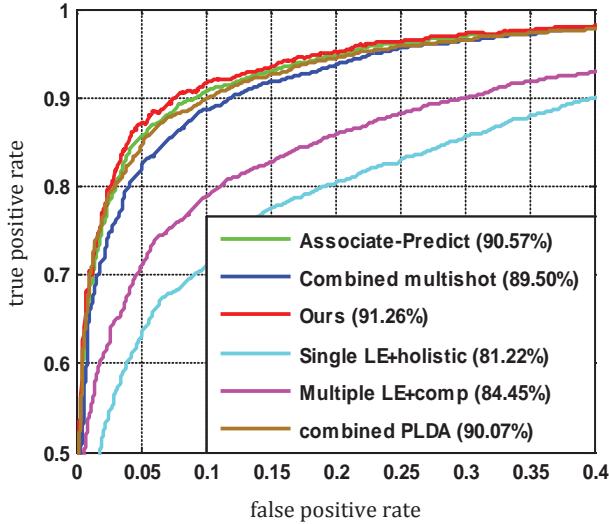


Figure 6. Verification performance on LFW with the outside training data.

detected locations of the eyes and mouth instead of their accurate alignment and warp, the accuracy in [3] is 90.47%. Since the in-face and out-face networks are constructed from patches of the aligned face images, the more accurate face alignment will lead to the more accurate face-patch networks. So it can also be predicted that the performance of our RP measure will be improved if performed on such accurately aligned faces.

5. Conclusion

This paper has proposed a random path (RP) measure for face recognition based on the path similarity defined from random walk in the network. To adopt the RP measure for face recognition, we construct two types of networks on face data: the in-face network and the out-face network. The in-face and out-face networks describe the local and global structural information of faces, respectively. We combine them to improve the recognition performance. Extensive experiments on the Multi-PIE and LFW face databases validate that the proposed RP measure has the superiority of discriminating complex faces with the large intra-personal variations including pose, illumination, and expression. This study has only examined the random path measure for face recognition. Our future work will explore the applications of the RP measure for other recognition tasks, such as image retrieval and object recognition.

References

- [1] T. Ahonen, A. Hadid, and M. Pietikäinen. Face recognition with local binary patterns. In *ECCV*, 2004.
- [2] P. N. Belhumeur, J. P. Hespanha, and D. J. Kriegman. Eigenfaces vs. fisherfaces: Recognition using class specific linear projection. *TPAMI*, 1997.
- [3] T. Berg and P. N. Belhumeur. Tom-vs-pete classifiers and identity-preserving alignment for face verification. In *BMVC*, 2012.
- [4] Z. Cao, Q. Yin, X. Tang, and J. Sun. Face recognition with learning-based descriptor. In *CVPR*, 2010.
- [5] C.-C. Chang and C.-J. Lin. Libsvm: a library for support vector machines. *TIST*, 2011.
- [6] D. Cox and N. Pinto. Beyond simple features: A large-scale feature search approach to unconstrained face recognition. In *FG*, 2011.
- [7] N. Dalal and B. Triggs. Histograms of oriented gradients for human detection. *CVPR*, pages 886–893, 2005.
- [8] A. S. Georghiades, P. N. Belhumeur, and D. J. Kriegman. From few to many: Illumination cone models for face recognition under variable lighting and pose. *TPAMI*, 2001.
- [9] R. Gross, I. Matthews, J. Cohn, T. Kanade, and S. Baker. Multi-pie. In *FG*, 2008.
- [10] G. Hua and A. Akbarzadeh. A robust elastic and partial matching metric for face recognition. In *ICCV*, 2009.
- [11] G. B. Huang, M. Mattar, T. Berg, E. Learned-Miller, et al. Labeled faces in the wild: A database for studying face recognition in unconstrained environments. In *Workshop on Faces in 'Real-Life' Images: Detection, Alignment, and Recognition*, 2008.
- [12] J. Kandola, J. Shawe-Taylor, and N. Cristianini. Learning semantic similarity. *Advances in neural information processing systems*, 2002.
- [13] L. Katz. A new status index derived from sociometric index. *Psychometrika*, 1953.
- [14] N. Kumar, A. C. Berg, P. N. Belhumeur, and S. K. Nayar. Attribute and simile classifiers for face verification. In *ICCV*, 2009.
- [15] N. K. L. Wiskott, J. Fellous and C. von der Malsburg. Face recognition by elastic bunch graph matching. *IEEE TPAMI*, 19(7), 1997.
- [16] Z. Li, D. Lin, and X. Tang. Nonparametric discriminant analysis for face recognition. *TPAMI*, 31(4), 2009.
- [17] M. Newman. *Networks: an introduction*. 2009.
- [18] T. Ojala, M. Pietikäinen, and T. Maenpää. Multiresolution gray-scale and rotation invariant texture classification with local binary patterns. *IEEE TPAMI*, 24(7):971–987, 2002.
- [19] N. Pinto, J. J. DiCarlo, and D. D. Cox. How far can you get with a modern face recognition test set using only simple features? In *CVPR*, 2009.
- [20] S. J. Prince and J. H. Elder. Tied factor analysis for face recognition across large pose changes. In *BMVC*, 2006.
- [21] S. T. Roweis and L. K. Saul. Nonlinear dimensionality reduction by locally linear embedding. *Science*, 2000.
- [22] L. K. Saul and S. T. Roweis. Think globally, fit locally: unsupervised learning of low dimensional manifolds. *The Journal of Machine Learning Research*, 2003.
- [23] H. S. Seung and D. D. Lee. The manifold ways of perception. *Science*, 2000.
- [24] T. Sim, S. Baker, and M. Bsat. The cmu pose, illumination, and expression database. *TPAMI*, 2003.
- [25] N. Sudha et al. Robust hausdorff distance measure for face recognition. *Pattern Recognition*, 2007.
- [26] J. Tenenbaum, V. Silva, and J. Langford. A global geometric framework for nonlinear dimensionality reduction. *Science*, 290:2319–2323, 2000.
- [27] J. B. Tenenbaum, V. De Silva, and J. C. Langford. A global geometric framework for nonlinear dimensionality reduction. *Science*, 2000.
- [28] M. A. Turk and A. P. Pentland. Face recognition using eigenfaces. In *CVPR*, 1991.
- [29] R. Wang, S. Shan, X. Chen, and W. Gao. Manifold-manifold distance with application to face recognition based on image set. In *CVPR*, 2008.
- [30] X. Wang and X. Tang. A unified framework for subspace face recognition. *TPAMI*, 26(9), 2004.
- [31] X. Wang and X. Tang. Random sampling for subspace face recognition. *IJCV*, 70(1), 2006.
- [32] L. Wiskott, J.-M. Fellous, N. Kuiger, and C. von der Malsburg. Face recognition by elastic bunch graph matching. *TPAMI*, 1997.
- [33] L. Wolf, T. Hassner, and Y. Taigman. Descriptor based methods in the wild. In *Faces in Real-Life Images Workshop in ECCV*, 2008.
- [34] Q. Yin, X. Tang, and J. Sun. An associate-predict model for face recognition. In *CVPR*, 2011.
- [35] Z. Zhu, P. Luo, X. Wang, and X. Tang. Deep learning identity-preserving face space. *ICCV*, 2013.