Neurocomputing 99 (2013) 339–346

Contents lists available at SciVerse ScienceDirect

Neurocomputing

journal homepage: www.elsevier.com/locate/neucom

Face identification using reference-based features with message passing model

Wei Shen [a], Bo Wang [b], Yueming Wang [a], Xiang Bai [a,*], Longin Jan Latecki [c]

[a] Huazhong University of Science and Technology, Wuhan, China
[b] University of Toronto, Toronto, Canada
[c] Temple University, Philadelphia, USA

ARTICLE INFO

Article history:
Received 19 August 2011
Received in revised form
4 June 2012
Accepted 9 June 2012
Communicated by D. Tao
Available online 10 August 2012

Keywords:
Face identification
Message passing model
One-shot similarity
Similarity space

ABSTRACT

In this paper, we propose a system for face identification. Given two query face images, our task is to tell whether or not they are of the same person. The main contribution of this paper comes from two aspects: (1) We adopt the one-shot similarity kernel [35] for learning the similarity of two face images. The learned similarity measures are then used to map a face image to reference images. (2) We propose a graph-based method for selecting an optimal set of reference images. Instead of directly working on the image features, we use the learned similarity to the reference images as the new features and compute the corresponding matching score of the two query images. Our approach is effective and easy to implement. We show encouraging and favorable results on the "Labeled Faces in the Wild" – a challenging data set of faces.

1. Introduction

Major progress has been made in the area of face recognition [26,53,54] and recognizing the face of a person from general images has become increasingly robust [32]. However, to be practical enough to deal with a range of variations such as lighting, pose change, and occlusion, a face recognition system often requires a number of training images from each person, which is not always available. In this paper, we tackle a somewhat different problem: face identification. Given a pair of face images, the system answers the question whether or not they are from the same person.

As shown in Fig. 1, in general, face identification has two main steps: (1) feature extraction and (2) feature-based matching (comparison). In the past, a variety of features have been proposed. Local Binary Pattern (LBP) [24] and its varieties [34] are among popular facial features, which represent each pixel as a binary number by thresholding its intensity value between its surrounding neighborhood's and encode these binary numbers by a histogram. Gabor wavelets [22] have also been used to describe the face appearance which normalizes and replaces the intensity values by edge responses [3]. SIFT descriptor [21] has been widely used for image representation due to its scale, orientation, and affine distortion invariant. Guillaumin et al. [14] proposed to use SIFT descriptors computed at the landmarks on the face (corners of the mouth, eyes, and nose) as the face feature. As reported in [34,15], the baseline results using these features by standard similarity measure, e.g., Euclidean distance, are not satisfactory. The reason is faces of the same person may have large variation caused by difference in lighting [2], pose [30], appearance, expression [11], partial occlusion [28] and cluster as illustrated in Fig. 2.

To increase the accuracy, many researchers focus on designing a more faithful similarity measure [5,9,10,19]. Most of them learn a Mahalanobis metric [37,33] based on an objective function which makes the distance between the data with the same label much smaller than the distances between those with different labels. Formally, the Mahalanobis distance between two data $x,y \in \mathbb{R}^d$ is $MD(x,y) = (x-y)^T\mathbf{M}(x-y)$, the goal is to learn a proper symmetric positive definite matrix $\mathbf{M} \in \mathbb{R}^{d\times d}$. Information theoretic metric learning (ITML) [6] is one of the state-of-the-art methods for Mahalanobis metric learning which uses an information theoretic approach to optimize $\mathbf{M}$ under the constraints that the similarity between each pair labeled "same" is below a specified threshold and the one between each pair labeled "different" is above another specified threshold. Chechik et al. [42] learnt a parametric similarity function which gives supervision on the *relative similarity* between two pairs of images through a bilinear form. Guillaumin et al. proposed a logistic

* Corresponding author.
E-mail addresses: shenwei1231@gmail.com (W. Shen), wangbo.yunze@gmail.com (B. Wang), youlihanshu@gmail.com (Y. Wang), xiang.bai@gmail.com (X. Bai), latecki@temple.edu (L.J. Latecki).

0925-2312/$ - see front matter © 2012 Elsevier B.V. All rights reserved.
http://dx.doi.org/10.1016/j.neucom.2012.06.032

discriminant-based metric learning method (LDML) [15], which relates to the classical logistic regression having the advantage of giving well-calibrated probabilities that pairs labeled "same". However, these metric learning methods face the problem of over-fitting.

In this paper, we proposed a new reference-based method to address the problems mentioned above. Rather than directly computing the distances of two face images, we utilize the similarities to learned reference faces as measurement. Directly measuring the distance between pairs of faces is quite difficult, because the variations between the images of the same face due to illumination and viewing direction are almost always larger than image variations due to change in face identify [44]. Therefore, by learning a proper set of face images and similarity functions, we hope the new features would be able to capture the intrinsic representation of a person's face. This paper includes two main contributions. First, we propose a new graph model for automatic selection of reference faces. A semi-supervised message passing model is used to select the references which are informative and representative. Messages (based on similarity) are propagated between faces and potential references so that we can achieve the most informative references. Second, we give a proper way to translate appearance feature to reference-based feature which captures better within-class similarity. The translation is based on one-shot similarity kernel, which is the average prediction of two classifiers learned from one example in a pair repeatedly. We show that the new similarity measure can make the distance space more compact and discriminative than metric learning approaches. We evaluate our approach on a challenging data set and obtain encouraging results.

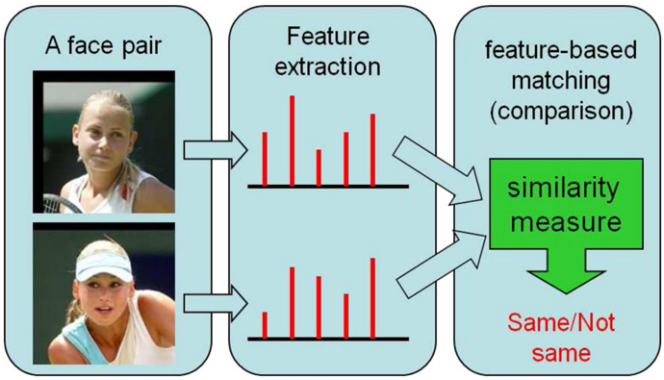


Fig. 1. The general framework of face identification. Any facial features and similarity measures can be employed in this framework.

The rest of this paper is organized as follows. Section 2 reviews some previous methods for face identification. Section 3.1 proposes the automatical algorithm for reference faces selection. In Section 3.2, we show how to identify faces on the similarity space. Section 4 provides our experimental results on the challenge data set "Labeled Faces in the Wild". Finally, the conclusions are given in Section 5.

2. Related work

In recent years, several approaches have been proposed for face identification. For example, Nowak et al. [38] quantize the corresponding patch pairs sampled from a face image pair by extremely randomized trees and build a similarity measure from the encoded patch pairs to compute the similarity of two face images. Hua et al. [43] take a part based face representation and propose to utilize both elastic and partial matching to handle the different visual variations in face images. Pinto et al. [27] using several element-wise differences (e.g., squared and absolute-value) between the feature vectors to represent a pair of face

Fig. 2. Face images of the same person may have large visual variations due to lighting (upper left), pose (upper right), facial appearance and expression (the middle row), partial occlusion (lower left) and cluster (lower right). The pairs surrounded by green rectangles and those surrounded by red rectangles are correctly and incorrectly classified by our method, respectively. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

images and boost the identification performance by multi-kernel learning. Both of the approaches in [15,39] are based on metric learning, the difference is the former one defines the similarity between a pair images by sigmoid function, while the later one uses cosine function. Representing an image by a reference set has received much attention during the last several years. The "simile" recognizer in [23] represents the face image by the scores of a sequence of classifiers. There are two main differences between our method and [23]: (1) In [23], they collect a reference set of labeled face images (identities are known) and about 600 faces are required for each reference person for training, whereas our method automatically selects the references from a set of unlabeled face images. (2) To train the classifier, they need to select several useful parts (e.g., eyebrows, eyes) from the face, which means their method also depends on part detectors. However, we directly compute the similarities between pairs of holistic faces. Wolf et al. [36] use a rank vector to represent an image which is generated by retrieving the image in the reference set. Cao et al. [4] densely sample the local feature vectors from the reference face images and generate the codebook by clustering the local feature vectors. Then the input face image is represented by the code quantized by the codebook.

3. Method

In this section, we propose our reference-based method for the problem of face identification. There are two main steps in our method: First, a proper set of reference faces is selected from the training set through the proposed constrained message passing model. Then, based on the learned references, we map the original features of two test faces to the feature space adopting the one-shot similarity kernel [35] and obtain two new feature vectors, which are used for the final identification. The discriminative power of the new feature vectors will depend on the selection of reference faces. We adopt a constrained message passing model to select reference faces, which is introduced below.

3.1. Constrained message passing model

In this section, we show how to choose the reference faces. A direct approach would choose the entire training set. However, this would introduce redundancy and large overhead in computing.

Most techniques for identifying references, e.g., k-means clustering, require both a pre-determined number of exemplars and an initial set of candidate exemplars. However, detecting references ("exemplars") goes beyond simple clustering as references themselves carry key information [41]. The optimal set of references is the one for which the sum of similarities of within-class and between-class points are minimized and maximized, respectively. Graph-based learning has been proved to be effective to improve the similarity measure between data points [56,55]. Here we also propose a graph model to learn a set of representative reference faces. This method is in spirit similar to message passing model (MPM) [41], but with some field-specific constraints.

We formulate the problem of finding the reference faces as a MAP inference problem by representing it using the binary grid factor-graph [12]. The graph model is shown in Fig. 3. We adopt an off-line approach to search the references among the training samples. Considering that there are N faces in the training set, we have $N \times N$ pairwise similarities $s(i,j)$ between N faces $i,j \in (1 \ldots N)$. We define N^2 hidden binary variables c_{ij}. Setting $c_{ij}=1$ denotes that face i's favorite reference is j, and $c_{ii}=1$ means that i is a reference face. The graphical model in Fig. 3 has nodes

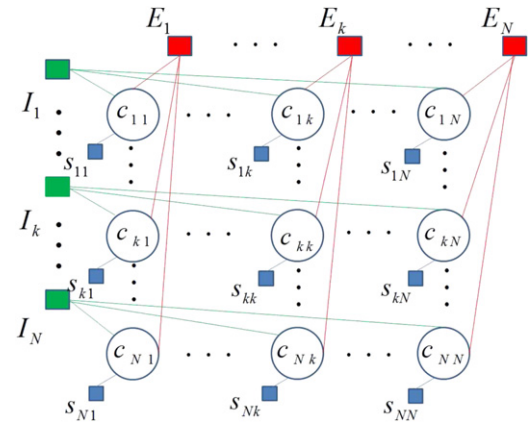


Fig. 3. A graph model for the binary variable factor graph. I node represents the constraint that one person can only choose one exemplar; E node emphasizes the restriction that a person must choose itself as an exemplar if others choose him as an exemplar. S nodes show that the latent variables c_{ij} is decided by the similarity s_{ij}.

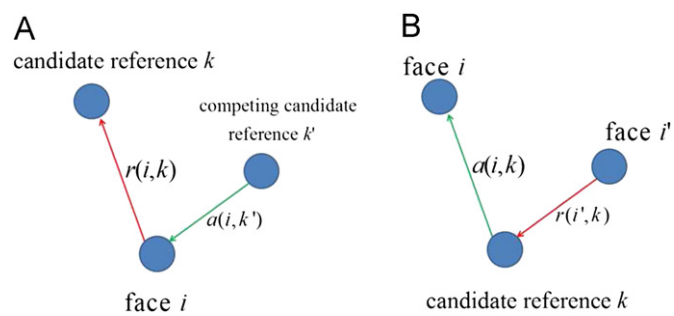


Fig. 4. The message passing model defines messages that are propagated between face points indicating the similarity each face has to another. (A) shows the propagation process from one face to all the other candidate exemplars. (B) shows the propagation process from one candidate exemplar to other faces. Note that all the information (r and a) are based on the similarity between all the faces.

representing three types of functions. Functions of type I ensure that each point can be assigned to at most one reference. Functions of type E introduce a consistency constraint: if $\exists i, c_{ij}=1 \Rightarrow c_{jj}=1$. Finally, the S_{ij} function nodes incorporate the input similarities s_{ij} between face points and evaluate to the similarity s_{ij} when $c_{ij}=1$. Formally, we define the functions as

$$I_i(c_{i1},c_{i2},\ldots,c_{iN}) = \begin{cases} -\infty & \text{if } \sum_j c_{ij} \neq 1, \\ 0 & \text{otherwise,} \end{cases} \tag{1}$$

$$E_j(c_{1j},c_{2j},\ldots,c_{Nj}) = \begin{cases} -\infty & \text{if } c_{jj}=0,\ \sum_i c_{ij} > 0, \\ 0 & \text{otherwise,} \end{cases} \tag{2}$$

$$S_{ij}(c_{ij}) = \begin{cases} s(i,j) & \text{if } c_{ij}=1, \\ 0 & \text{otherwise.} \end{cases} \tag{3}$$

Hence the MAP formulation of the message passing model in our framework is

$$S(c_{11},\ldots,c_{NN}) = \sum_{i,j} S_{ij}(c_{ij}) + \sum_i I_i(c_{i:}) + \sum_j E_j(c_{:j}), \tag{4}$$

where $c_{i:}$ represents $\{c_{i1},c_{i2},\ldots,c_{iN}\}$ and $c_{:j}$ represents $\{c_{1j},c_{2j},\ldots,c_{Nj}\}$.

The approximate MAP setting for the c_{ij} variables is inferred by the max-sum algorithm, a log-domain equivalent of max-product [40]. Fig. 4 illustrates that two kinds of messages are passed

between data points including candidate references and normal faces. Candidate references are those who have potential to be the references, while the normal faces are the remaining ones. The messages have an intuitive interpretation: the "responsibility" $r(i,j)$ indicates how much face point i wants the other face j to be its exemplar; the "availability" $a(i,j)$ is an indicator of the extent face i considers itself as a perfect exemplar of face j. The messages are updated and exchanged between the hidden variables and all the function nodes in the following way [41]:

$$a(i,j)=\begin{cases}\sum_{k\neq j}\max[0,r(k,j)], & i=j,\\ \min\left[0,r(j,j)+\sum_{k\neq j,i}\max[0,r(k,j)]\right], & i\neq j,\end{cases} \tag{5}$$

$$r(i,j)=s(i,j)-\max_{k\neq j}(s(i,k)+a(i,k)). \tag{6}$$

To apply the factor graph model to our problem of finding the reference faces, we must add some constraints. If two faces are known to be the same face, they should be linked together. To ensure the constraint, the method in [13] added some fictitious "meta-faces" or *MTF*s. The *MTF*s allow them to explicitly enforce the must-link constraints and cannot-link constraints, as well as to propagate must-link constraints and construct a mechanism for cannot-link constraints to be propagated. However, this approach significantly complicates the updating of messages. Therefore, we adopt a simpler strategy. For those who should be linked together, denoted as $SL=\{(i,j)|i \text{ should linked to } j\}$ with assuming that $(i,i)\in SL$, we can assign a face i to a reference by changing the updating Eq. (5) to

$$a(i,j)=\begin{cases}\sum_{k\neq j}\max[0,r(k,j)], & (i,j)\in SL,\\ \min\left[0,r(j,j)+\sum_{k\neq j,i}\max[0,r(k,j)]\right], & (i,j)\notin SL.\end{cases} \tag{7}$$

After convergence we are ready to select the reference faces. Each value $a(i,i)+r(i,i)$ indicates the confidence for face i to be a reference. Then, we set a threshold to choose the final references as

$$References=\{i|a(i,i)+r(i,i)>\theta\}. \tag{8}$$

By changing the threshold $\theta\geq 0$, we can control the number of references. In our experiments, we set $\theta=0$.

We stress that the proposed method does not fix the number of references. Through passing messages (similarities), one face tells every other face its ranked list of favorite reference. A candidate reference tells other faces the degree of compatibility to be a reference. Every sent message is evaluated through a simple computation on the basis of the received messages and the similarity matrix. After several message-passing rounds, every face knows its favorite reference. In Section 4.2.3, we show that MPM is a powerful technology to select the proper references and to automatically determine their number.

3.2. Identification in the similarity space

In this section, we show how to measure the similarity between a pair of test faces based on the reference faces.

We represent face images as d-dimensional vectors defined on the original feature space $\mathbb{R}^d$. Let $\Upsilon=\{\gamma_1,\gamma_2,\ldots,\gamma_n\}$ be the feature vectors of the reference faces and $\{x_i,x_j\}$ be the ones of two test faces. Given a similarity function $f(\cdot,\cdot)$, the face image $x\in\mathbb{R}^d$ is mapped into the similarity space [25] $\mathbb{R}^n$. Then, the feature vector of a test face x can be described by an n-dimensional similarity vector $S(x,\Upsilon)=[f(x,\gamma_1),f(x,\gamma_2),\ldots f(x,\gamma_n)]^T$. Note that, each element of $S(x,\Upsilon)$ describes a similarity measure between a reference face and the test face. Intuitively, if x_i is similar to x_j, $S(x_i,\Upsilon)$ should be similar to $S(x_j,\Upsilon)$. This property has been discussed explicitly in [7].

Clearly, the key problem for obtaining S is to find a proper similarity function $f(\cdot,\cdot)$. Euclidean metric is the most frequently used similarity measure, the similarity between a novel face x and a reference face $\gamma\in\Upsilon$ based on Euclidean metric could be defined by a redial basis function (RBF) kernel [8] as

$$f(x,\gamma)=\exp\left(-\frac{ED(x,\gamma)}{\sigma^2}\right), \tag{9}$$

where $ED(\cdot,\cdot)$ is the Euclidean distance and σ is a hyper-parameter. However, Euclidean metric is not faithful in the original feature space. Here we utilize the One-Shot (OS) similarity [35] measure to map the original feature space into similarity space.

To describe our method better, we briefly review OS similarity here. To compute the OS similarity for two feature vectors x_i and x_j, a set **A** of "negative" training feature vectors is required which have different labels (identifies) from those we wish to compare. First a discriminative model is learned by taking **A** as the "negative" set and x_i as a single positive datum. This model is then used to classify x_j, and obtain a classification prediction score Score1. Then the above process is repeated with the roles of x_i and x_j switched to obtain Score2. The OS similarity depends on the classifier used, such as SVM or LDA, it can be a signed value which gives us a measure of how likely one vector to be compared is to belong to the same class as the other. Note that, the OS similarity is not a distance. It is a similarity measure between $-\infty$ and $+\infty$ where high positive value means similar and low value (negative) means dissimilar. In our experiments, the OS similarities are predominantly negative.

Following the previous work [4,20], to pursue the compactness, we apply a dimensionality reduction technique and then L_2 normalization to the similarity vectors $S(x,\Upsilon)$. Finally, we use the obtained compressed similarity vector $S_c(x,\Upsilon)$ for identification. Hence, the similarity of a pair $\{x_i,x_j\}$ can be measured by the standard metrics between the compressed similarity vectors $ED(S_c(x_i,\Upsilon),S_c(x_j,\Upsilon))$. We use Euclidean metric as the default metric on the similarity space without any especial statement. We refer to this method as RBFM, for reference-based feature mapping.

In our experiments, to reduce the computing complexity, we employ LDA within the OS scheme. Since the negative set **A** is used repeatedly, and the positive class, which contains just one element, does not contribute to the within class covariance matrix, we can compute the within class covariance matrix off-line and the OS similarity can be efficiently computed on-line.

Table 1
Baseline results on LFW.

Feature	Euclidean	Hellinger	χ^2
LBP	70.78% ± 0.7	70.52% ± 0.5	68.48% ± 0.4
SIFT	69.42% ± 0.5	69.78% ± 0.5	70.03% ± 0.7
Gist	68.13% ± 0.6	69.12% ± 0.5	69.18% ± 0.6
LLC	71.18% ± 0.3	69.92% ± 0.3	70.25% ± 0.2

Table 2
Our results on LFW obtained by using a single feature.

Feature	PCA	KPCA	MDS	PCA(CS)
LBP	78.72% ± 0.5	78.78% ± 0.5	78.60% ± 0.5	77.32% ± 0.3
SIFT	80.20% ± 0.3	81.02% ± 0.3	80.30% ± 0.3	80.83% ± 0.3

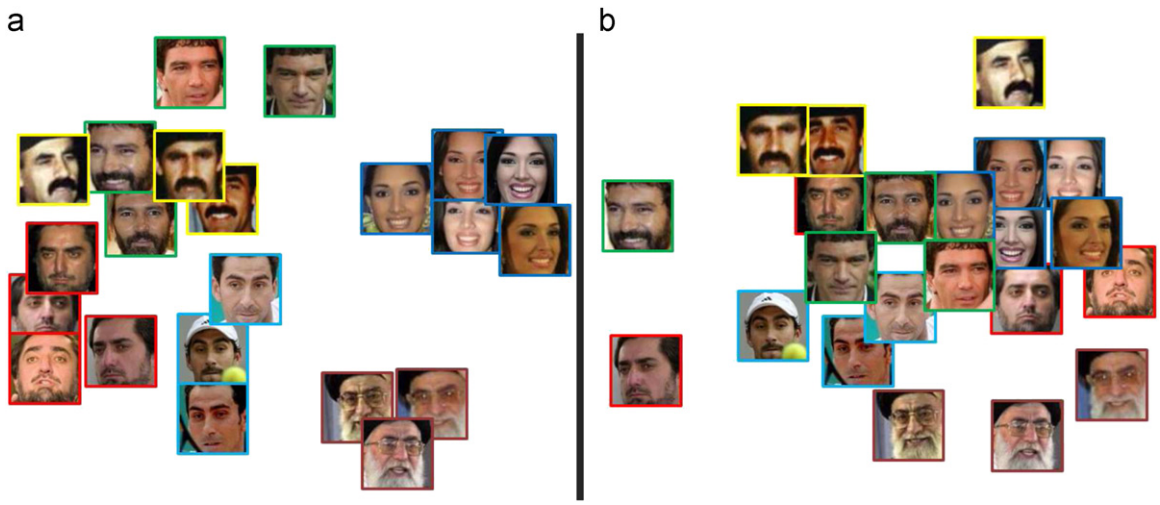


Fig. 5. (Left) Visualization results after the dimension reduction of the similarity vectors vs. (right) the original feature vectors. Images positioned according to pairwise Euclidean distances between dimension reduction results of our similarity vectors (left) and those between original feature vectors (right). Color frames encode subject identities. The faces in (a) are clustered better than (b). (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

4. Experimental results

4.1. Data set

The *Labeled Faces in the Wild* (LFW)[1] [18] is a benchmark data set for face identification, which contains 13,233 face images collected from *Yahoo! News* in 2002–2003. Each face has been labeled with the name of the person pictured. In total 5749 people appear in the images, 1680 of them have two or more distinct photos [16]. LFW is a very challenging data set since the faces in it show a big variety in lighting, pose, appearance, etc. as shown in Fig. 2. To test identification algorithm, the data set provides 10 independent folds for cross validation and each fold contains 600 face pairs with half labeled "same" and half labeled "different". In turn, nine folds are chosen as training data and the remaining one is used for testing. Note that, when predicting the label of a test pair, neither the identities of the people in the training set nor the information from the other testing pairs can be used. The identification accuracy is the percentage of the pairs identified correctly. The final identification result is reported as the average accuracy of the 10 runs. To test our method on LFW, at each time, we use one of the nine training folds to produce the negative set and one of the eight remaining folds to generate the reference faces. LFW has two versions, one is the original version and the other is the funneled version, in which images are aligned by the method in [16]. Following [34,15,4,27,17,23], we only use the funneled version in our experiments.

4.2. Identification result on LFW

4.2.1. Comparison to the baseline

The baseline results on LFW are reported in Table 1, which are obtained by applying three standard metrics (Euclidean, Hellinger and χ^2 distance) to three basic features (LBP [24], SIFT [15], Gist [45]) and a bag-of-features (BoF) representation LLC [46]. Note that, all distances and features lead to a comparable accuracy of 68% and 71%. Since LBP and SIFT are the two most popular features used for face identification [34,15,39], for fair comparison, we only use these two basic features hereafter.

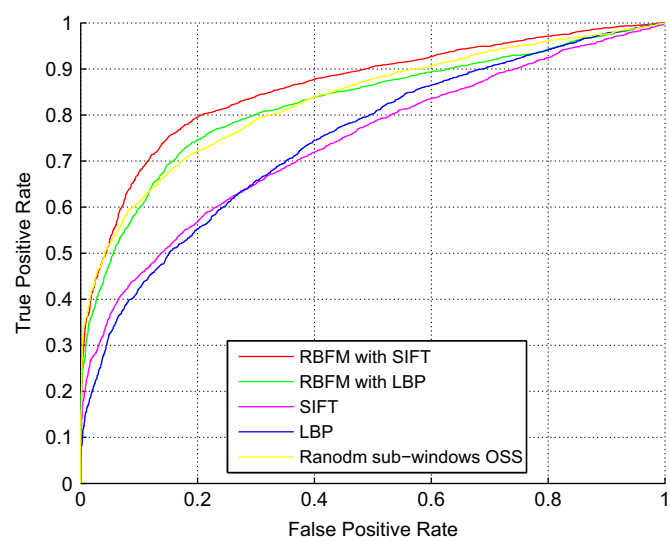


Fig. 6. ROC curve comparison with the two basic features and the random sub-windows based representation using OS similarity [35].

Table 2 summarizes the performance of our method by using only a single feature; encouragingly, all our results significantly outperform the baseline. In our experiments, about 340 faces are selected by MPM from the training set as the references. To perform dimensionality reduction, a large number of approaches can be adopted, such as [47–52,29,1]. Here, we tried three different dimensionality reduction techniques in our method RBFM, including two linear dimensionality reduction methods PCA and MDS [1] and one nonlinear dimensionality reduction method Kernel PCA (KPCA) [29]. We take the implementation from the Matlab Toolbox for dimensionality reduction[2] and we use about 240 dimensions (the total variance explained by the principal components is equal or greater than 90%) for the next computation. Unlike the method in [4], these dimensionality reduction techniques only give a small improvement in our method. Besides Euclidian distance, we could also use the

[1] The dataset can be downloaded from http://vis-www.cs.umass.edu/lfw/.

[2] http://homepage.tudelft.nl/19j49/Matlab_Toolbox_for_Dimensionality_Reduction.html

weighted cosine similarity [39] as the similarity measure on the similarity space. The weights are related to the eigenvalues obtained in PCA process. We list the results obtained by weighted cosine similarity in the last column of Table 2. The best performance achieved is 81.02% ± 0.3, using SIFT and KPCA with Gaussian kernel. To give the intuitive explanation, we visualize pairwise Euclidean distances between the results after the dimensionality reduction of our similarity vectors and compare with those between the original feature vectors in Fig. 5. We randomly select six individuals from the LFW set having at least three images each. The selected face images are then positioned on the plane by computing the 2D Multidimensional-Scaling of their distances (MATLAB's `mdscale` function). Fig. 6 shows the ROC curve comparison between the two basic features and RBFM based on them with KPCA. The method in [35] is also based on OS similarity. Hence we also include its ROC curve in Fig. 6 for comparison. Note that, our ROC curve of RBFM based on LBP is comparable with the ROC curve of the method in [35] and the one based on SIFT is better than it.

4.2.2. Quantitative evaluation of the identification methods based on a single feature

For further comparison, we list the results obtained by the methods based on only a single feature in Table 3. Note that the "OS Similarity" in [34] is computed directly on the holistic faces as we do in our experiments. In [35] each face image pair is represented by randomly selecting 1000 image coordinate and sampling patches of normally distributed sizes. The image pair is represented by a vector containing 1000 OS similarities, one for every corresponding patch pair. Our best result for LBP is 78.78% ± 0.5, better than the other two results related to OS similarity and the highest score for a single feature obtained by our method is not significantly different from the score reported in [4], which is the best for a single feature.

Table 3
Performance comparison with the methods using a single feature.

Method	Feature	Accuracy
OS similarity [34]	LBP	74.63% ± 0.5
OS similarity, patches [35]	LBP	76.37% ± 0.7
Single LE [4]	Single LE	81.22% ± 0.5
LDML [15]	SIFT	77.50% ± 0.5
RBFM	SIFT	**81.02% ± 0.3**

Table 4
Accuracy on LFW based on different references selection methods.

Method	PCA	KPCA	MDS
k-means	76.20% ± 0.4	76.81% ± 0.5	76.62% ± 0.4
Random	75.82% ± 0.5	76.04% ± 0.5	75.93% ± 0.5
MPM	78.72% ± 0.5	78.78% ± 0.5	78.60% ± 0.5

4.2.3. Quantitative evaluation of the methods for references selection

To quantitatively show the power of constrained message passing model in our method, we compare it to the reference faces selected by other methods such as *k*-means clustering and random selection. The results on LFW are shown in Table 4. All the results are computed based on the same single feature LBP. Clearly, no matter which dimensionality reduction method is used, constrained message passing model achieves better performance. We also show and compare the reference face images selected by MPM and *k*-means in Fig. 7. We pick the first 16 reference face images from the first folds of LFW. The reference faces images selected by MPM come from more individuals and have larger variations in pose, appearance and expression, so they are more informative.

4.2.4. Comparison to the state-of-the-art

As pointed out in [34,15,4], different features and metrics may have complementary information, thus the identification accuracy can be improved by combining them. Following the previous works, we linearly combine two features with three different methods of dimension reduction (*cf.* Table 2), totally eight scores. The distances are transformed into similarities by Eq. (9). The linear combinations are learnt for each fold independently. In the following, we refer to these combined methods as Multi-RBFM. To compare with the state-of-the-art, we list previously published results on LFW in Table 5 and plot the ROC curves in Fig. 8. All reported results are obtained by combining several features or methods. It needs to be mentioned that the method in [31] uses additional information such as pose. Wolf et al.'s work [36] and Nguyen et al.'s work [39] adopt the face alignment algorithm in [36] other than the funneled version. The method in [23] uses extra data to train a sequence of trait classifiers which is outside the LFW test protocol. Therefore, it is not fair to compare with these methods. Under the LFW test protocol, without any extra data, the performance of our method outperforms others on the funneled version, which is a remarkable achievement.

Table 5
Performance comparison to the previously published methods on LFW.

Method	Accuracy
Hybrid descriptor-based [34]	78.47% ± 0.5
V1-like/MKL [27]	79.35% ± 0.6
Multi-shot [31]	83.98% ± 0.4
Combined b/g samples [36]	86.83% ± 0.3
Attribute and simile classifiers [23]	85.29% ± 1.2
Multiple LE [4]	84.45% ± 0.5
MERL+Nowak [17]	76.18% ± 0.6
Combined-LDML [15]	79.27% ± 0.6
CSML [39]	88.00% ± 0.4
Multi-RBFM	**84.50% ± 0.4**

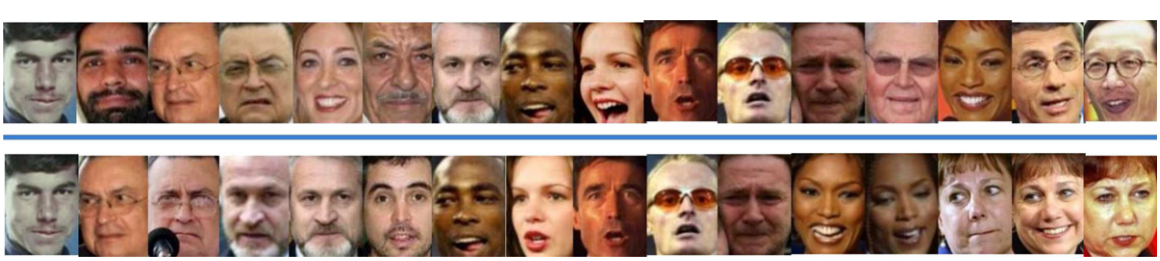


Fig. 7. Reference faces selected by MPM (Top) and *k*-means (Bottom), respectively.

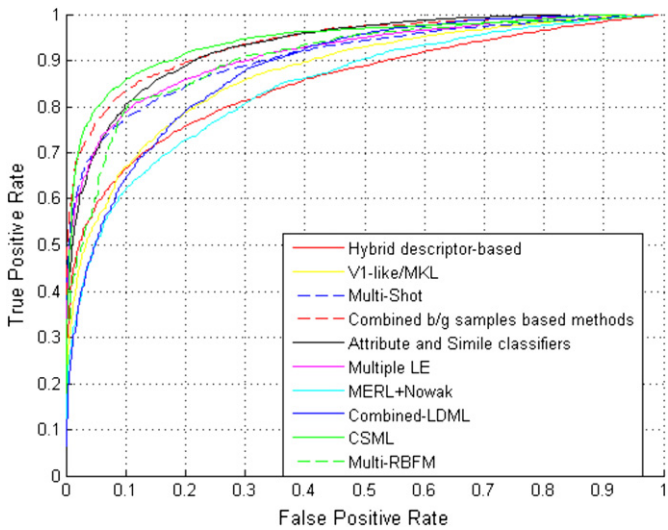


Fig. 8. ROC curve comparison to the state-of-the-art on LFW.

5. Conclusions and future work

We propose a framework for face identification problem in which faces are mapped to a new feature space determined by their similarities to reference faces. A graph-based model named constrained message passing model is proposed to select an optimal set of reference faces. The presented experimental results on LFW data set show that the proposed method is comparable to the state-of-the-art methods. This demonstrates the discriminative power of our reference-based features. Our future work includes three main aspects: (1) improvement of this approach by learning the different distributions of similarities among same-class and different-class; (2) pose and illumination estimation for more accurate face representation and matching; (3) extension to other problems such as image retrieval and image annotation.

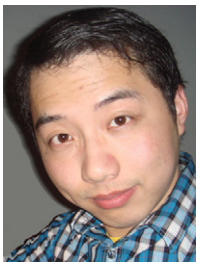
Acknowledgments

This work was supported by the Fundamental Research Funds for the Central Universities' HUST 2011TS110 and National Natural Science Foundation of China #60903096.

References

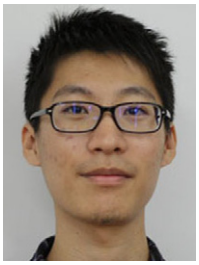
[1] A. Bronstein, M. Bronstein, R. Kimmel. Generalized multidimensional scaling: a framework for isometry-invariant partial surface matching, Proc. Natl. Acad. Sci. USA 103 (5) (2006) 1168–1172.
[2] P.N. Belhumeur, J.P. Hespanha, D.J. Kriegman, Eigenfaces vs. fisherfaces: recognition using class specific linear projection, IEEE Trans. Pattern Anal. Mach. Intell. 19 (7) (1997) 711–720.
[3] R. Brunelli, T. Poggio, Face recognition: features versus templates, in: IEEE International Conference on Computer Vision and Pattern Recognition, vol. 15, 1993, pp. 1042–1052.
[4] Z. Cao, Q. Yin, X. Tang, J. Sun, Face recognition with learning-based descriptor, in: IEEE International Conference on Computer Vision and Pattern Recognition, 2010.
[5] S. Chopra, R. Hadsell, Y. LeCun, Learning a similarity metric discriminatively, with application to face verification, in: IEEE International Conference on Computer Vision and Pattern Recognition, 2005.
[6] J. Davis, B. Kulis, P. Jain, S. Sra, I. Dhillon, Information theoretic metric learning, in: Proceedings of the ICML, 2007.
[7] S. Edelman, Representation and Recognition in Vision, MIT Press, 1999.
[8] M.D. Buhmann, Radial Basis Functions: Theory and Implementations, Cambridge University Press, 2003.
[9] A. Ferencz, E. Learned-Miller, J. Malik, Building a classification cascade for visual identification from one example, in: Proceedings of the ICCV, 2005.
[10] A. Ferencz, E. Learned-Miller, J. Malik, Learning to locate informative features for visual identification, Int. J. Comput. Vis. 77 (2008) 3–24.
[11] A.S. Georghiades, P.N. Belhumeur, D.J. Kriegman. From few to many: illumination cone models for face recognition under variable lighting and pose, in: IEEE International Conference on Computer Vision and Pattern Recognition, 2007.
[12] I. Givoni, B. Frey, A binary variable model for affinity propagation, Neural Comput. 21 (6) (2009) 1589–1600.
[13] I. Givoni, B. Frey. Semi-supervised affinity propagation with instance-level constraints, in: Proceedings of the AISTATS, 2009.
[14] M. Guillaumin, T. Mensink, J. Verbeek, C. Schmid, Automatic face naming with caption-based supervision, Int. J. Pattern Recognition Artif. Intell. 20 (8) (2006) 117–138.
[15] M. Guillaumin, J. Verbeek, C. Schmid, Is that you? metric learning approaches for face identification, in: Proceedings of the ICCV, 2009.
[16] G. Huang, V. Jain, E. Learned-Miller, Unsupervised joint alignment of complex images, in: Proceedings of the ICCV, 2007.
[17] G. Huang, M. Jones, E. Learned-Miller, LFW results using a combined NOWAK plus MERL recognizer, in: Real-Life Images Workshop at ECCV, 2008.
[18] G. Huang, M. Ramesh, T. Berg, E. Learned-Miller, Labeled Faces in the Wild: a Database for Studying Face Recognition in Unconstrained Environments, Technical Report, 07-49, University of Massachusetts, Amherst, 2007.
[19] V. Jain, A. Ferencz, E. Learned-Miller, Discriminative training of hyper-feature models for object identification, in: Proceedings of the BMVC, 2006.
[20] Y. Ke, R. Sukthankar, PCA-sift: a more distinctive representation for local image descriptors, in: Proceedings of the CVPR, 2004.
[21] D. Lowe, Distinctive image features from scale-invariant keypoints, Int. J. Comput. Vis. 60 (2) (2004) 91–110.
[22] E. Meyers, L. Wolf, Using biologically inspired features for face processing, Int. J. Comput. Vis. 76 (2008) 93–104.
[23] N. Kumar, A. Berg, P. Belhumeur, S. Nayar, Attribute and simile classifiers for face verification, in: Proceedings of the ICCV, 2009.
[24] T. Ojala, M. Pietikäinen, T. Mäenpää, Multiresolution gray-scale and rotation invariant texture classification with local binary patterns, IEEE Trans. Pattern Anal. Mach. Intell. 24 (2002) 971–987.
[25] E. Pekalska, P. Paclík, R. Duin, A generalized kernel approach to dissimilarity based classification, J. Mach. Learn. Res. 2 (2002) 175–211.
[26] P. Phillips, P. Flynn, T. Scruggs, K. Bowyer, J. Chang, K. Hoffman, J. Marques, J. Min, W. Worek, Overview of the face recognition grand challenge, in: Proceedings of the CVPR, 2005.
[27] N. Pinto, J. DiCarlo, D. Cox, How far can you get with a modern face recognition test set using only simple features? in: Proceedings of the CVPR, 2009.
[28] F. Samaria, A. Harter, Parameterization of a stochastic model for human face identification, in: Proceedings of the WACV, 1994.
[29] B. Scholkopf, A. Smola, K. Müller, Nonlinear component analysis as a kernel eigenvalue problem, IEEE Trans. Pattern Anal. Mach. Intell. 10 (5) (1998) 1299–1319.
[30] T. Sim, S. Baker, M. Bsat, The CMU pose, illumination and expression database, IEEE Trans. Pattern Anal. Mach. Intell. 25 (12) (2003) 1615–1618.
[31] Y. Taigman, L. Wolf, T. Hassner, Multiple one-shots for utilizing class label information, in: Proceedings of the BMVC, 2009.
[32] A. Wagner, J. Wright, A. Ganesh, Z. Zhou, Y. Ma, Towards a practical face recognition system: robust alignment and illumination by sparse representation, in: Proceedings of the CVPR, 2009.
[33] K. Weinberger, J. Blitzer, L. Saul, Distance metric learning for large margin nearest neighbor classification, in: Proceedings of the NIPS, 2006.
[34] L. Wolf, T. Hassner, Y. Taigman, Descriptor Based Methods in the Wild, Real-Life Images Workshop at ECCV, 2008.
[35] L. Wolf, T. Hassner, Y. Taigman, The one-shot similarity kernel, in: Proceedings of the ICCV, 2009.
[36] L. Wolf, T. Hassner, Y. Taigman. Similarity scores based on background samples, in: Proceedings of the ACCV, 2009.
[37] E. Xing, A. Ng, M. Jordan, S. Russell, Distance metric learning, with application to clustering with side information, in: Proceedings of the NIPS, 2004.
[38] E. Nowak, F. Jurie, Learning visual similarity measures for comparing never seen objects, in: Proceedings of the CVPR, 2007.
[39] H.V. Nguyen, L. Bai, Cosine similarity metric learning for face verification, in: Proceedings of the ACCV, 2010.
[40] F. Kschischang, B. Frey, P.-A. Duin, Factor graphs and the sum-project algorithm, IEEE Trans. Inf. Theory 47 (2) (2001).
[41] B. Frey, D. Dueck, Clustering by passing messages between data points, Science 315 (2007) 927–976.
[42] G. Chechik, V. Sharma, U. Shalit, S. Bengio, An online algorithm for large scale image similarity learning, in: Proceedings of the NIPS, 2009.
[43] G. Hua, A. Akbarzadeh, A robust elastic and partial matching metric for face recognition, in: Proceedings of the ICCV, 2009.
[44] Y. Moses, Y. Adini, S. Ullman, Face recognition: the problem of compensating for changes in illumination direction, in: Proceedings of the ECCV, 1994.
[45] A. Oliva, A. Torralba, Modeling the shape of the scene: a holistic representation of the spatial envelope, Int. J. Comput. Vis. 42 (3) (2001) 145–175.
[46] J. Wang, J. Yang, K. Yu, F. Lv, T. Huang, Y. Gong, Locality-constrained linear coding for image classification, in: Proceedings of the CVPR, 2010.
[47] T. Zhou, D. Tao, X. Wu, Manifold elastic net. A unified framework for sparse dimension reduction, Data Min. Knowl. Discov. 22 (3) (2011) 340–371.
[48] N. Guan, D. Tao, Z. Luo, B. Yuan, Manifold regularized discriminative nonnegative matrix factorization with fast gradient descent, IEEE Trans. Image Process. 20 (7) (2011) 2030–2048.

[49] N. Guan, D. Tao, Z. Luo, B. Yuan, Non-negative patch alignment framework, IEEE Trans. Neural Networks 22 (8) (2011) 1218–1230.
[50] W. Bian, D. Tao, Max-min distance analysis by using sequential SDP relaxation for dimension reduction, IEEE Trans. Pattern Anal. Mach. Intell. 33 (5) (2011) 1037–1050.
[51] D. Tao, X. Li, X. Wu, S.J. Maybank, General tensor discriminant analysis and Gabor features for gait recognition, IEEE Trans. Pattern Anal. Mach. Intell. 29 (10) (2007) 1700–1715.
[52] D. Tao, X. Li, X. Wu, S.J. Maybank, Geometric mean for subspace selection, IEEE Trans. Pattern Anal. Mach. Intell. 31 (2) (2009) 260–274.
[53] Y. Xua, W. Zuo, Z. Fan, Supervised sparse representation method with a heuristic strategy and face recognition experiments, Neurocomputing 79 (1) (2012) 125–131.
[54] H. Li, Y. Wei, L. Li, Y. Yuan, Similarity learning for object recognition based on derived kernel, Neurocomputing 83 (15) (2012) 110–120.
[55] J. Yu, D. Tao, M. Wang, Adaptive hypergraph learning and its application in image classification, IEEE Trans. Image Process. 21 (7) (2012) 3262–3272.
[56] J. Yu, D. Liu, D. Tao, H.-S. Seah, Complex object correspondence construction in 2D animation, IEEE Trans. Image Process. 20 (11) (2011) 3257–3269.



Wei Shen received his B.S. degree in Electronics and Information Engineering from the Huazhong University of Science and Technology (HUST), Wuhan, China, in 2007. Currently, he is a Ph.D. candidate at HUST and work in Microsoft Research Asia as an intern.

Bo Wang received his B.S. degree in Electronics and Information Engineering from the Huazhong University of Science and Technology (HUST), Wuhan, China, in 2010. Currently, he is a master student at University of Toronto, Toronto.



Yueming Wang received his B.S. degree in Electronics and Information Engineering Department, Huazhong University of Science and Technology (HUST), Wuhan, China, in 2011. Now, he works at Media and Communication Lab, HUST as a graduate student. His interest includes pattern recognition and image processing.

Xiang Bai received his B.S. and M.S. degree both in electronics and information engineering from Huazhong University of Science and Technology (HUST), Wuhan, China, in 2003 and in 2005, respectively. He obtained his Ph.D. degree from HUST in 2010. From January 2006 to May 2007, he worked in the Department of Computer Science and Information, Temple University. From October 2007 to October 2008, he worked in the University of California, Los Angeles as a joint Ph.D. student. Now he is a faculty of EI Department, HUST. His research interests include computer graphics, computer vision, and pattern recognition.

Longin Jan Latecki is an associate professor in the Department of Computer and Information Sciences at the Temple University in Philadelphia. He is the winner of the 25th Pattern Recognition Society Award together with Azriel Rosenfeld for the best paper published in the journal Pattern Recognition in 1998. He received the main annual award from the German Society for Pattern Recognition (DAGM), the 2000 Olympus Prize. He is a member of the Editor Board of Pattern Recognition and chairs the IS&T/SPIE annual conference series on Vision Geometry. He has published and edited over 150 research articles and books. His main research areas are shape representation and shape similarity, object recognition, robot mapping, video analysis, data mining, and digital geometry.