

Research Article

A Novel Approach to Face Verification Based on Second-Order Face-Pair Representation

Qiang Hua, Chunru Dong, and Feng Zhang 

Key Laboratory of Machine Learning and Computational Intelligence of Hebei Province,
College of Mathematics and Information Science, Hebei University, Baoding 071002, China

Correspondence should be addressed to Feng Zhang; amyfzhang@yahoo.com

Received 24 February 2018; Accepted 15 April 2018; Published 25 June 2018

Academic Editor: Zhile Yang

Copyright © 2018 Qiang Hua et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Face representation and matching are two essential issues in face verification task. Various approaches have been proposed focusing on these two issues. However, few of them addressed the joint optimal solutions of these two issues in a unified framework. In this paper, we present a second-order face representation method for face pair and a unified face verification framework, in which the feature extractors and the subsequent binary classification model design can be selected flexibly. Our contributions can be summarized in the following aspects. First, a novel face-pair representation method that employs the second-order statistical property of the face pairs is proposed, which retains more information compared to the existing methods. Second, a flexible binary classification model, which differs from the conventionally used metric learning, is constructed based on the new face-pair representation. Finally, we verify that our proposed face-pair representation can benefit from large training datasets. All the experiments are carried out on Labeled Face in the Wild (LFW) to verify the algorithm's effectiveness against challenging uncontrolled conditions.

1. Introduction

Face recognition has been extensively studied in the field of computer vision and pattern recognition, which in general can be categorized into two tasks, that is, face verification and face identification. In this paper, we focus on the former one. Face verification normally works on the data of face pairs; that is, given a face pair, we need to decide whether they are from the same person or not. However, this is not an easy task due to various challenges, such as position, background, pose, lighting, and occlusion (Figure 1). Like most pattern recognition systems, face verification task has two key components, that is, face representation and face matching. These two components are the central parts that most of the researchers are concerned about. Although miscellaneous algorithms have been proposed focusing on these two components, few of them considered both in a unified framework. Intuitively, jointly formulating these two components may result in a promising performance.

Over the years, many face representation approaches have been proposed. According to the intrinsic characteristics that

are depicted, these can be classified into three simple categories: (1) local features descriptors, (2) holistic features descriptors, and (3) features descriptors based on deep learning.

Local features descriptors have been proven to be very effective for texture description. For instance, the local binary pattern (LBP) [1, 2] encodes the structure distribution into a histogram by computing the relative intensity magnitude difference between each pixel and its neighbors based on the predefined rules. Harr-like features [3] are rule-based local feature descriptors. However, such handcrafted encoding methods are suspected to get optimal encoding for a specific task. Alternatively, more flexible and uniformly distributed local descriptors can be learned for face recognition task [4]. HoG [5] and SIFT features [6] can also be categorized into local features descriptors. And SIFT also considers robustness to scale variations by building the scale spaces. Local features descriptors are more stable to local changes such as illumination, expression, and inaccurate alignment. Gabor wavelets [7–9] captured the local structure corresponding to specific spatial frequency (scale), spatial locality, and selective



FIGURE 1: Confusing face pairs selected from LFW [25]. (a) shows images from different people, while (b) shows the opposite case.

orientation. They have been demonstrated to be discriminative and robust to illumination and expression changes.

The holistic features descriptors mainly focus on feature transformation, which consists of various algorithms that create new but fewer features while having higher discriminatory power in a different space than the original feature space. This can also be used for feature reduction. In such transformation process, the advantages of statistical techniques are employed. The typical techniques are PCA [10, 11], LDA [12], ICA [13], LPP [14], and so forth. These techniques are categorized as subspace learning methods and are distinguished by their ability to preserve some desired properties of the data. The holistic feature represents the visual features of the whole face. In general, the efficiency of holistic features relies heavily on well aligned face, that is, uniform in scale, pose, illumination, and so forth. Consequently, holistic feature is more sensitive to misalignment [15] and other variations from intrinsic changes (e.g., expression, pose) and ambient environment changes (e.g., side lighting).

Deep learning [16–18] has received increasing interest in computer vision and face verification recently, and a number of deep learning methods have been proposed in the literature. One merit of deep learning methods is that they preserve the face image’s neighborhood relations and spatial locality in their latent higher-level feature representations. However, the deep learning methods need a huge number of labeled samples, which brings about the computation assumption issues.

Face matching is the other important component of face verification. Usually, a metric is necessary for measuring the similarity of an input face pair, and then a learned threshold is used to perform the matching task. Considering the optimal similarity (or distance) metric with regard to the specific feature space, lots of automatic metric learning algorithms have been proposed under different objectives. For instance, Guillaumin et al. [19] learned a Mahalanobis-like metric using the logistic discriminant method with the objective that positive pairs have smaller distances than negative pairs; Davis et al. formulated the Mahalanobis learning problem as

that of minimizing the differential relative entropy between two multivariate Gaussians, while adding an additional regularization term that emphasizes a prior of the covariance matrix in their objective [20]. There are some other metric learning methods which are different in their objective for the specific tasks in the literature [21–23] as well, while they share the same framework of face verification that is depicted in Figure 2.

Mahalanobis-like distance learning can also be viewed as an Euclidean distance in the linearly transformed feature space. This refers to a more extensively researched topic, that is, kinds of projection techniques, which we have described in the above section as a feature representation approach. However, all these projections are independent of the subsequent classifiers. Reconsidering the face verification task, if we can represent a face pair using a unified vector, then the face verification task can subsequently be treated as a simple binary classification problem. That is, we determine whether face pairs are mismatched (i.e., the different persons) or matched (i.e., the same person). The representation not only is able to describe the dissimilarities between face pairs, but also naturally implies a simile classifier [24]. Therefore, this data representation strategy endows great flexibility for the subsequent classification model design.

Motivated by this idea, classification-related data organization for face pairs is proposed, which is more flexible and valid to the following classification model design. And consequently, a joint optimal approach to put face representation and face matching into a unified framework is provided. This paper is organized as follows. Section 2 briefly reviews the fundamental knowledge, which is used to introduce the motivation of the proposed approach. In Section 3, a novel representation of a face pair, which employs the second-order statistical property, is proposed. And a framework of a face verification system based on the proposed face-pair representation is proposed in Section 3 too. Three basic experiments under the framework are implemented to illustrate the efficiency of the proposed method in Section 4. All the experiments are carried out on Labeled Face in the Wild.

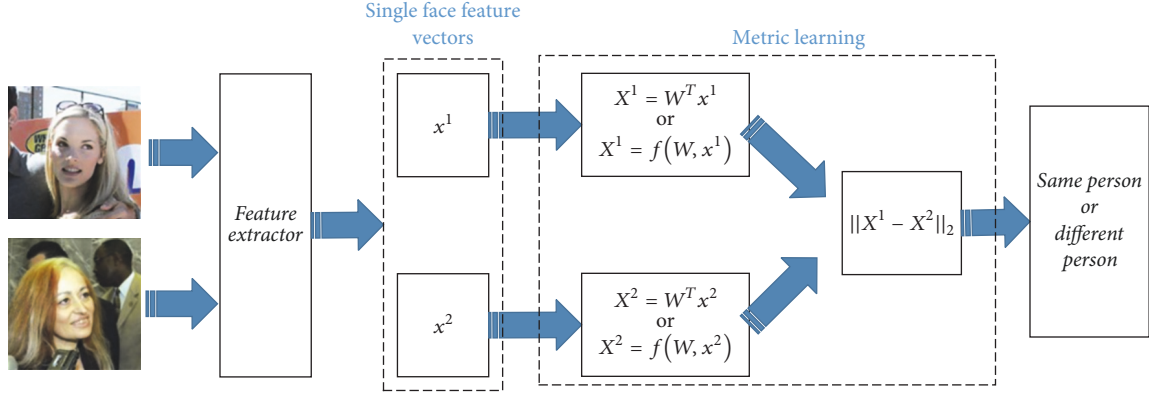


FIGURE 2: The overall framework of face verification based on conventional Mahalanobis distance learning.

Finally, a conclusion is presented at the end of the paper in Section 5.

2. Preliminaries

In this section, we will briefly review the conventional Mahalanobis distance metric learning and some classical binary classifiers.

2.1. Mahalanobis Distance Metric Learning. Metric learning algorithms have been proposed and some of them have already been applied to tackle the problem of face verification over the decades. The common objective of these methods is to learn a suitable metric, which can be used to reduce the distance between positive face pairs and enlarge that of negative pairs. Since the metric can be parameterized using a positive semidefinite matrix M , the metric learning can be transformed into learning the positive semidefinite matrix M .

Let $X = \{(x^1, x^2, \dots, x^n) \in \mathbb{R}^{d \times n}$ be a training set of n samples, where $x^i \in \mathbb{R}^d$ is the i th sample and n is the total number of training samples. The task of the traditional Mahalanobis distance metric learning is to seek a square matrix $M \in \mathbb{R}^{d \times d}$. Based on the matrix M , we can define the squared distance between two samples using the following equation:

$$d_M(x^i, x^j) = (x^i - x^j)^T M (x^i - x^j), \quad (1)$$

where $x^i \in \mathbb{R}^d$ is the i th sample in the training set and $M \in \mathbb{R}^{d \times d}$ is a symmetric and positive semidefinite matrix. It can be decomposed as follows:

$$M = L^T L, \quad (2)$$

where $L \in \mathbb{R}^{p \times d}$ performs a linear transformation and normally $p \ll d$.

Therefore, the squared distance defined in (1) can be reformulated by

$$\begin{aligned} d_M(x^i, x^j) &= (x^i - x^j)^T M (x^i - x^j) \\ &= (x^i - x^j)^T L^T L (x^i - x^j) \\ &= \|L(x^i - x^j)\|_2^2. \end{aligned} \quad (3)$$

Equation (3) shows that learning a Mahalanobis metric is equivalent to learning a linear transformation L , which actually projects an original sample into a low dimensional subspace due to the fact that $p \ll d$. The Mahalanobis metric in the original space is degenerated into an Euclidean distance between two samples in the transformed space.

2.2. Logistic Regression. Consider a set of m training samples, denoted by $\{x^i, y^i\}_{i=1}^n$, where $x^i \in \mathbb{R}^d$ is the i th sample and $y^i \in \{-1, +1\}$ denotes its corresponding class label. The likelihood function of these n samples is defined by $\prod_{i=1}^n \text{Prob}(y^i | x^i)$, and the average logistic loss is defined as [26]

$$\begin{aligned} f(w, c) &= -\frac{1}{n} \log \prod_{i=1}^n \text{Prob}(y^i | x^i) \\ &= \frac{1}{n} \sum_{i=1}^n \log(1 + \exp(-y^i (w^T x^i + c))). \end{aligned} \quad (4)$$

Therefore, we can determine the parameters w and c by minimizing the average logistic loss; that is,

$$\min f(w, c) = \frac{1}{n} \sum_{i=1}^n \log(1 + \exp(-y^i (w^T x^i + c))). \quad (5)$$

The minimization problem will lead to a smooth convex optimization problem because the average logistic loss function is a smooth and convex function.

2.3. Support Vector Machine. Given training samples $x^i \in \mathbb{R}^d$ ($i = 1, 2, \dots, m$) in two classes and a label vector $y^i \in \{-1, +1\}$ denoting the corresponding label of samples x^i , the

Support Vector Machine can determine the weight of the samples by solving the following optimization problem [27]:

$$\begin{aligned} \min_{\omega, b, \xi_i} \quad & \frac{1}{2} \omega^T \omega + C \sum_{i=1}^n \xi_i \\ \text{subject to} \quad & y^i (\omega^T \phi(x^i) + b) \geq 1 - \xi_i, \\ & \xi_i \geq 0, \quad i = 1, 2, \dots, n, \end{aligned} \quad (6)$$

where $\phi(x^i)$ maps the sample x^i into a higher-dimensional space and $C > 0$ is the regularization parameter. Considering the dimensionality of the vector variable ω , we usually solve its dual problem:

$$\begin{aligned} \min_{\alpha} \quad & \frac{1}{2} \alpha^T M \alpha - e^T \alpha \\ \text{subject to} \quad & y^i \alpha = 0, \\ & 0 \leq \alpha_i \leq C, \quad i = 1, 2, \dots, n, \end{aligned} \quad (7)$$

where $e = [1, 1, \dots, 1]^T$ is the vector of all ones and M is a positive semidefinite matrix, with the element $m_{ij} = y^i y^j K(x^i, x^j)$, where $K(x^i, x^j) = \phi(x^i)^T \phi(x^j)$ is called the kernel function.

3. Face Verification Based on Second-Order Face-Pair Representation

In this section, a novel face-pair representation approach is proposed. We firstly present a second-order face-pair representation, and therefore the task of face verification can be viewed as a binary classification problem in a unified framework. On the one hand, the second-order representation should keep more discriminant information; therefore, we expect that such novel viewpoint of face verification problem can benefit from large training samples. On the other hand, since the proposed unified framework integrates feature extraction and classifier selection for the face verification problem, it should enhance the flexibility of the system.

3.1. Reformulate Face Verification. Face verification is a problem of determining whether two face images depict the same person or not. Its formal description is as follows.

Given two faces, denoted by x^1 and x^2 , the task of face verification is to decide whether these two faces (a face pair) are from different persons or the same one.

The conventional solution to the face verification is formulated as a metric learning problem, for example, learning a Mahalanobis distance between x^1 and x^2 , which is employed to present the similarity of a face pair. The final results are obtained by checking the predicted distance based on a properly chosen threshold, which actually implements a matching procedure.

However, there are four shortcomings in the metric-learning-based face verification process: (1) the positive definiteness of M is required; (2) it is difficult to obtain a closed solution to M ; (3) some prior domain knowledge is

difficult to be embedded flexibly; (4) it is hard to choose the appropriate threshold to determine the final result.

To overcome these drawbacks, we propose a unified face verification framework, in which the face-pair representation is employed instead of a single-face representation. Based on the face-pair representation, the face verification problem can be transformed into a binary classification problem defined on a face-pair measurement space. In order to distinguish it from single-face representation, we refer to this face-pair representation.

For a pair of faces, denoted by x^1 and x^2 , the measurement between a face pair is defined as

$$X_k = d(x_k^1, x_k^2), \quad (8)$$

where $k = 1, 2, \dots, n$. Then, the face verification problem can be transformed into binary classification problem on a face-pair measurement space.

As a result, the target of the face verification is equivalent to training a function f , which is defined on the transformed face-pair measurement space X :

$$f(X_k) = f(d(x_k^1, x_k^2)), \quad (9)$$

where $k = 1, 2, \dots, n$, $f(X_k) \in \{0, 1\}$ functions as a binary classifier, 0 denotes that the face pair X_k is from different persons, and 1 denotes that the face pair X_k is from the same person.

Specifically, the classifier is degenerated into a linear classifier when we obtain a linear function.

$$f(X_k) = W^T X_k, \quad (10)$$

where W corresponds to the weight to be learned in a binary classification problem.

3.2. A Second-Order Face-Pair Representation. In the reformulated binary classification task of face verification, it is essential to construct a representation space for face pairs. First, we define positive samples and negative samples, respectively, as face pairs from the same person and those from different persons. We intend to find a unified representation of a face pair X_k , which is different from the conventional single-face representation. Intuitively, the representation should keep more discriminant information so as to facilitate the following verification task.

Since face verification aims to find the similarity between two faces in the given face pairs, we can simply represent the face pair as the difference between the measures of two faces. Kumar et al. [24] utilized the absolute value of the difference between two trait vectors and the weighted product was defined to measure the similarity of two faces. However, the definition is given subjectively, and less similarity information is retained.

It is well known that the Mahalanobis distance computes the distance between two samples, while taking into account the covariance structure across the n -dimensional features.

Motivated by this characteristic, we rewrite the initial Mahalanobis distance defined in Section 2 as follows:

$$\begin{aligned}
 d_M(x^1, x^2) &= (x^1 - x^2)^T M (x^1 - x^2) \\
 &= \text{tr} \left((x^1 - x^2)^T M (x^1 - x^2) \right) \\
 &= \text{tr} \left(M (x^1 - x^2) (x^1 - x^2)^T \right) \\
 &= \text{tr} \left((x^1 - x^2) (x^1 - x^2)^T M \right) \\
 &= (M_{(\cdot)})^T \left((x^1 - x^2) (x^1 - x^2)^T \right)_{(\cdot)},
 \end{aligned} \tag{11}$$

where (\cdot) denotes the operation that rearranges a matrix into a column vector.

If we denote $((x^1 - x^2)(x^1 - x^2)^T)_{(\cdot)}$ in (11) by X , that is,

$$\left((x^1 - x^2) (x^1 - x^2)^T \right)_{(\cdot)} \triangleq X, \tag{12}$$

then (11) can be rewritten as

$$\begin{aligned}
 d_M(x^1, x^2) &= (M_{(\cdot)})^T \left((x^1 - x^2) (x^1 - x^2)^T \right)_{(\cdot)} \\
 &= W^T X,
 \end{aligned} \tag{13}$$

where the weight W corresponds to $M_{(\cdot)}$; it can be learned by training a binary classifier. 0 denotes that the face pair X is from different persons and 1 denotes that the face pair X is from the same person.

Remark 1. The definition of X in (12) contains covariance-like feature representation; such second-order representation of X reserves more information for the face verification problem.

Remark 2. Due to the symmetry of the feature matrix for face-pair representation, we can use the upper triangle data only to reduce the storage and computation burden without any information loss.

The second-order feature representation has close relationship with Mahalanobis distance metric. When we choose the linear model as the binary classifier in the above architecture (Figure 3), then the face verification process is equivalent to computing a Mahalanobis-like distance between two faces, while face verification process can be implemented nonlinearly as well when the binary classifier is chosen as a nonlinear classifier. Throughout this paper, we will apply this mechanism for face-pair representation, while the basic representation of a single face can be various.

3.3. A Face Verification Framework Based on the Second-Order Face-Pair Representation. The proposed second-order face-pair representation transfers the face verification problem as a more general two-category classification task. The overview of face verification based on our proposed face-pair representation method is illustrated in Figure 4. The architecture can

be decomposed into three main processes: (1) basic feature extraction for a single face, which can be implemented using an appropriate feature extractor; (2) second-order face-pair representation based on the difference of two faces; (3) a more flexible two-category classifier.

It is worth noting that this is a general framework based on second-order face-pair representation. In practice, the basic feature extraction can be implemented with great flexibility while incorporating prior knowledge from the specific problem domain. For instance, a face can be encoded by global feature descriptors or local feature descriptors. The classification module can also be implemented with great flexibility. For instance, logistic regression can be employed to tackle the linear classification problem, while the kernel support vector machine is more suitable to nonlinear classification problem.

Another advantage of our proposed framework is that the system can benefit from learning with a large amount of data. Intuitively, most statistical-based learning methods will benefit from the large amount of training data. Because there is some second-order statistical information embedded in the proposed face-pair representation, the proposed method will make use of the large amount of training data more effectively. The experiment results in Section 4 do verify that. The dimension of the second-order face-pair representation grows in quadratic speed with the dimension of the single-face representation, which may limit its applications in high-dimensional data settings. That is, if the dimension of the feature vector of single face is d , then that of face-pair measure space will be $d(d+1)/2$. Therefore, a feature dimension reduction technique is necessary. Usually, dimension reduction techniques are exploited to tackle this problem. However, there is a limited compression ratio while attempting to preserve as much information as possible to facilitate the analysis of the latter classification process.

4. Implementation

In this section, experiments are conducted to verify the performance and effectiveness of our proposed method.

4.1. Dataset. We choose the Labeled Faces in the Wild [25] dataset as our testing bed, which is the de facto standard dataset for face verification. The dataset is highlighted by three characteristics: (1) enriched in both people number and image number per people; (2) challenges of variations in scale, pose, lighting, background, hairstyle, clothing, expression, color saturation, image resolution, focus, and so forth; (3) flexibility in data organization. Data can be used in both “restricted” and “unrestricted” settings. In “restricted” setting, data are provided in the form of paired faces. In “unrestricted” setting, a large amount of paired faces can be generated as needed. Following the standard performance reporting protocol, we report our performance on the 10-folder sets of view 2 [25]. These 10-folder sets are independently divided, that is, one person only in one set, 300 positive pairs, and 300 negative pairs for each set. While reporting performance, each one set is held out for testing and another

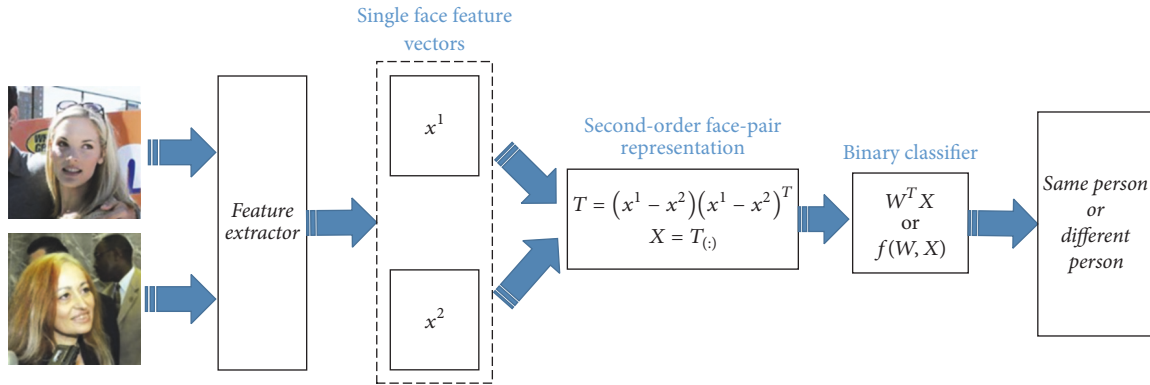


FIGURE 3: The overall framework of face verification based on the second-order face-pair representation.

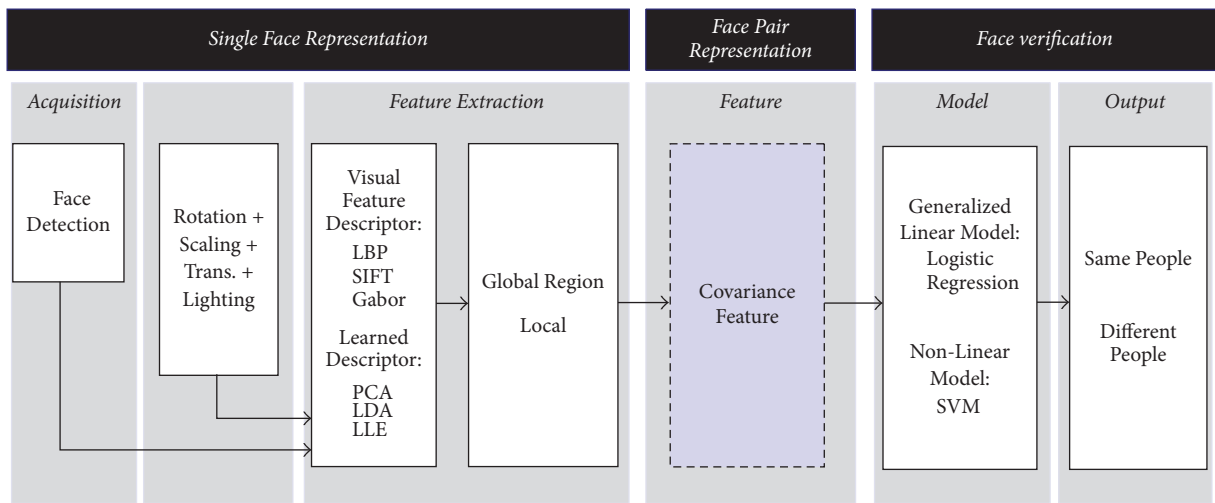


FIGURE 4: The proposed face verification framework based on the second-order feature representation.

9 sets are used for training. This will generate 10 sets of testing scores. The estimated averaged accuracy and standard deviation are computed following the standard definition in [25].

We use the aligned version of this dataset referred to as “funneled,” in which all the faces are globally aligned. Semantic facial parts are localized using an off-the-shelf facial feature detector [34], which outputs 9 points about corners of eyes, mouth, and nose. Based on these points, we induce another 3 points as the center of each eye and mouth. Then, we extract a 128-dimensional feature vector using SIFT descriptor at these 12 localized points with three patch sizes (16, 32, and 46) and three blur scales, leading to a $9 \times 128 = 1152$ -dimensional descriptor for each facial part.

4.2. Experimental Setup. The performance of a face verification system relies on many factors, such as basic feature representation, feature transformation, metric learning, and classification design. We try to fully consider the effects of these factors while evaluating the proposed method. The main consideration and the corresponding methods are summarized in several aspects below.

First, the issue of misalignment of face image will affect the performance of the proposed system. However, what we

focus on in this paper is the face-pair representation and its influence on the performance of face verification system. It is not necessary to make accurate global face alignment, since geometric alignment based on some specified facial parts location will not perform well with larger changes of pose and expression. What is more, facial parts detection is a hard job, and incorrect localization of one part will lead to great misalignment of a whole face.

Second, we consider the basic feature representation of a single face. Kinds of feature extraction methods can be utilized to represent a single face and each feature extraction method has its own advantages and deficiencies. Here, we employ SIFT feature [6] as the basic feature for single-face representation, considering its robustness against the common challenges, such as the variations of scale, lighting, and rotation, especially in real-life scenarios.

Third, local feature representation of a single face is useful in practical face verification systems, especially for enhancing the system’s robustness against great changes of pose, expression, and lighting, since incorrect localization of one local part will not affect other components. In this implementation, we use the facial feature detector proposed in [34] to extract several semantic local facial components, which

are assembled together to form the face, such as eyes, nose, mouth corners, and eyebrows.

The final process of the pipeline is classifier design. Most popular classifiers can be exploited to perform such two-category classification task. Besides this, the face identifier can also be constructed in two ways: one global classifier for a whole face and one classifier for each facial part. The popular techniques as logistic regression and support vector machine are chosen as our classification model. In this paper, we utilize the off-the-shelf implementation of these two techniques for our experiments, that is, SLEP [35] for logistic regression and libSVM [36] for SVM.

Based on these above settings, we design three experiments. The first one is used to illustrate the performance of our method with different feature descriptors (i.e., global features and local features) and with different classifiers (i.e., logic regression and kernel-SVM). The second experiment is implemented to illustrate that the verification accuracy of our approach can benefit from the growing number of training face pairs. Finally, a comparison between several face verification algorithms and our method is carried out in terms of the verification accuracy and ROC curves.

4.3. Experimental Results. We first evaluate the performance of the system using the proposed face-pair representation in a global way, where all the SIFT features extracted on these points are concatenated into a single vector as the single-face descriptor. Since the dimension of a single face is high, we apply PCA to obtain a more compact single-face representation. The dimension of features reduced by PCA is also an important issue; we experimentally chose 100 as the best trade-off between accuracy and complexity. However, in the global way, we find that if we choose 100 principle components, the preserved energy is only approximately 67%. Such low energy ratio implies that much information was lost in the dimension reduction process. In order to preserve more useful information within an endurable source cost, we train a local classifier for each facial point. The process is similar to the global way, and we also select the appropriate dimension for each local descriptor, and then we simply average all the local classifiers results to make the final decision. The experimental results are listed in Table 1.

Table 1 shows that when we choose logistic regression as the binary classifier, the performance of the system with local feature descriptors is enhanced compared to the one with global feature descriptors. Since different binary classifiers can be adopted as the classifier model in our proposed framework, we can also choose SVM with RBF kernel (KSVM) as the classification model; the performance of the system with local feature descriptors is enhanced from 80.35% to 83.07% compared to the one with global feature descriptors. We also noted that integrating the nonlinear classifier into the system will improve the performance compared to using a linear classifier, that is, logistic model. The corresponding ROC curve is shown in Figure 5.

Due to the second-order statistical property of the proposed face-pair representation method, the corresponding face verification model will benefit from a large amount of training data. A single-face encoding method [24] has

TABLE 1: Performance of the proposed method with different feature descriptors and classifiers.

Data organization manner	Classifier	Recog. rate (%)
Global	LR	76.10 \pm 0.67
Local	LR	78.12 \pm 0.47
Global	KSVM	80.35 \pm 0.64
Local	KSVM	83.07 \pm 0.52

TABLE 2: Recognition rates of the method used in [24] and performance of our method with large training data.

Methods	Recog. rate (%)
Abs. prod. [24]	80.79 \pm 0.75
Our method	82.74 \pm 0.63
Our method with large data	84.02 \pm 0.52

TABLE 3: Face verification accuracy of different methods in the LFW-restricted setting.

Method	Recog. rate (%)	AUC
Eigenfaces [28]	60.02 \pm 0.79	0.6483
Nowak [29]	73.93 \pm 0.49	0.8131
Hybrid descriptor-based [30]	78.47 \pm 0.51	0.8532
V1-like/MKL [31]	79.35 \pm 0.55	0.8742
APEM (fusion) [32]	84.08 \pm 1.20	0.8541
Spartans [33]	87.55 \pm 0.21	0.8982
Our method	86.08 \pm 0.62	0.8783

shown its efficiency in face verification problem, where only 73 attributes scores are used to encode each face. Based on this single-face encoding method, we get a 1701-dimensional vector for a face pair. In order to enlarge the size of the training dataset, we randomly add 1000 positive pairs and the same number of negative samples, to verify the trend of performance improvement with the number of training data. The final results show that improved performance is obtained with increased number of training samples. From Table 2, we can see that more than one percent point improvement is obtained compared to the trained model under the restricted setting, and nearly four percent point improvement is achieved compared to the initial abs. prod. [24].

Finally, a comparison between several face verification algorithms and our method is carried out in terms of the verification accuracy and ROC curves. The experiments are carefully performed with the protocol of Image-Restricted, No Outside Data. All experiments use centered 150×150 crops of ‘‘LFW-funneled’’ images. The recognition rate and AUC of different methods are shown in Table 3, and the corresponding ROC curve is shown in Figure 6. It shows that our method outperforms most of the existing methods except for Spartans, which preprocessed the data more delicately.

5. Conclusions

In this paper, we derived a second-order face-pair representation based on the formula of the conventional Mahalanobis

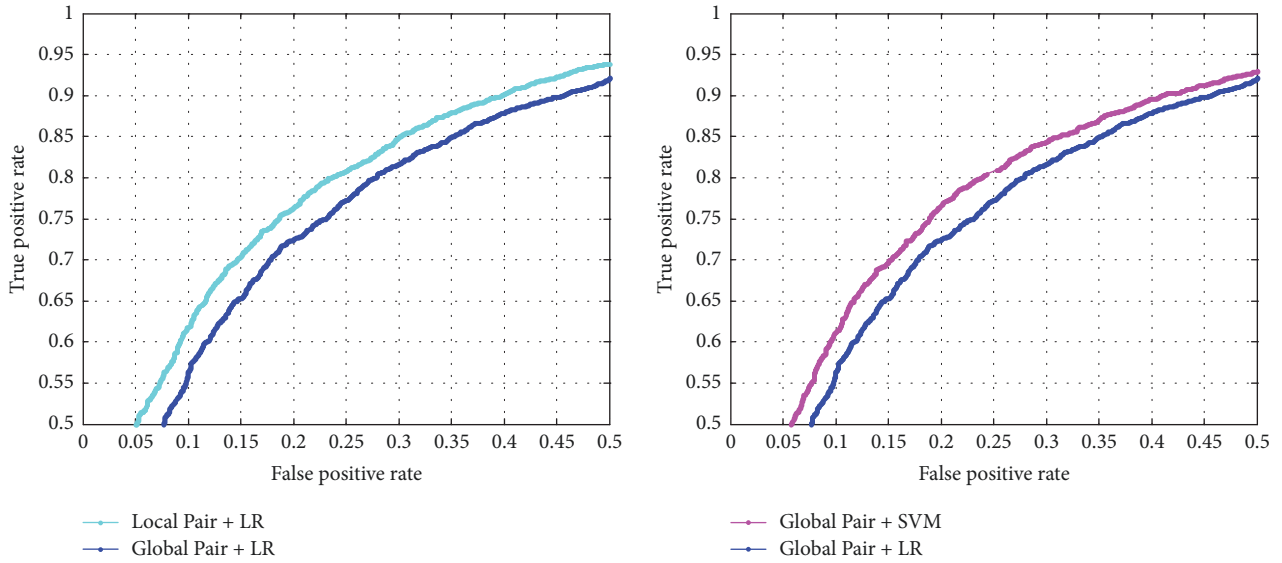


FIGURE 5: ROC curves of the system with different settings.

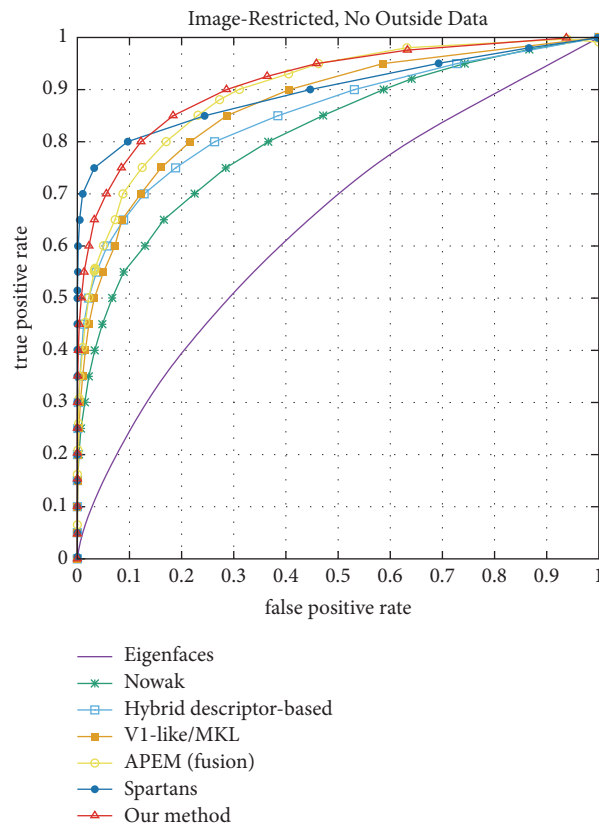


FIGURE 6: ROC curves of our method and typical techniques in LFW-restricted settings.

distance metric learning. This face-pair representation provides more discriminate information. Based on the face-pair representation, a unified face verification framework is put forward, which makes the proposed approach more flexible. The flexibility can be summarized in two aspects: (1) flexibility of choosing feature extractor for a single face

and (2) flexibility of selecting the classifiers. What is more, this novel face-pair representation brings the second-order statistical property of the difference between two faces. Three basic experiments on LFW have been implemented to illustrate the efficiency and validation of the proposed method.

Face verification is a complex problem that has substantial connections with deep convolutional neural networks, and their integration could improve the performance, both in theory and in application aspects. In our future work, we will try to integrate the deep CNN into the unified framework, to obtain more satisfied feature representation and improve the performance of face verification task.

Data Availability

The data used to support the findings of this study are available from the corresponding author upon request.

Conflicts of Interest

The authors declare that they have no conflicts of interest.

Acknowledgments

This research is supported by the Natural Science Foundation of Hebei Province (F2018201115, F2018201096) and the Youth Scientific Research Foundation of Education Department of Hebei Province (QN2015026, QN2017019).

References

- [1] T. Ojala, M. Pietikäinen, and T. Mäenpää, "Multiresolution gray-scale and rotation invariant texture classification with local binary patterns," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 24, no. 7, pp. 971–987, 2002.
- [2] T. Ahonen, A. Hadid, and M. Pietikäinen, "Face description with local binary patterns: application to face recognition," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 28, no. 12, pp. 2037–2041, 2006.
- [3] P. Viola and M. J. Jones, "Robust real-time face detection," *International Journal of Computer Vision*, vol. 57, no. 2, pp. 137–154, 2004.
- [4] Z. Cao, Q. Yin, X. Tang, and J. Sun, "Face recognition with learning-based descriptor," in *Proceedings of the 2010 IEEE Computer Society Conference on Computer Vision and Pattern Recognition, CVPR 2010*, pp. 2707–2714, USA, June 2010.
- [5] N. Dalal and B. Triggs, "Histograms of oriented gradients for human detection," in *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR '05)*, vol. 1, pp. 886–893, June 2005.
- [6] D. G. Lowe, "Distinctive image features from scale-invariant keypoints," *International Journal of Computer Vision*, vol. 60, no. 2, pp. 91–110, 2004.
- [7] L. Wiskott, J.-M. Fellous, N. Krüger, and C. von der Malsburg, "Face recognition by elastic bunch graph matching," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 19, no. 7, pp. 775–779, 1997.
- [8] B. Zhang, S. Shan, X. Chen, and W. Gao, "Histogram of Gabor phase patterns (HGPP): a novel object representation approach for face recognition," *IEEE Transactions on Image Processing*, vol. 16, no. 1, pp. 57–68, 2007.
- [9] Y. Pang, Y. Yuan, and X. Li, "Gabor-based region covariance matrices for face recognition," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 18, no. 7, pp. 989–993, 2008.
- [10] M. Turk and A. Pentland, "Eigenfaces for recognition," *Cognitive Neuroscience*, vol. 3, no. 1, pp. 71–86, 1991.
- [11] C. Liu, D. Zhou, Z. Wang, D. Yang, and G. Song, "Damage Detection of Refractory Based on Principle Component Analysis and Gaussian Mixture Model," *Complexity*, vol. 2018, pp. 1–9, 2018.
- [12] O. C. Hamsici and A. M. Martinez, "Bayes optimality in linear discriminant analysis," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 30, no. 4, pp. 647–657, 2008.
- [13] M. S. Bartlett, M. H. Lades, and T. J. Sejnowski, "Independent component representations for face recognition, Face Image Analysis by Unsupervised Learning," in *Face Image Analysis by Unsupervised Learning*, pp. 528–539, Springer US, 2001.
- [14] X. He and P. Niyogi, "Locality preserving projections," *Advances in Neural Information Processing Systems*, vol. 45, no. 1, pp. 186–197, 2005.
- [15] S. Shan, Y. Chang, W. Gao, B. Cao, and P. Yang, "Curse of Misalignment in face recognition: problem and a novel Mis-alignment learning solution," in *Proceedings of the 6th IEEE International Conference on Automatic Face and Gesture Recognition*, pp. 314–320, May 2004.
- [16] G. B. Huang, H. Lee, and E. Learned-Miller, "Learning hierarchical representations for face verification with convolutional deep belief networks," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pp. 2518–2525, 2012.
- [17] Y. Sun, Y. Chen, X. Wang, and X. Tang, "Deep learning face representation by joint identification-verification," *NIPS Proceedings*, pp. 1701–1708, 2014.
- [18] O. M. Parkhi, A. Vedaldi, and A. Zisserman, "Deep Face Recognition," in *Proceedings of the British Machine Vision Conference 2015*, pp. 41.1–41.12, Swansea.
- [19] M. Guillaumin, J. Verbeek, and C. Schmid, "Is that you? metric learning approaches for face verification," in *Proceedings of the IEEE International Conference on Computer Vision*, pp. 498–505, Kyoto, Japan, 2009.
- [20] J. V. Davis, B. Kulis, P. Jain, S. Sra, and I. S. Dhillon, "Information-theoretic metric learning," in *Proceedings of the 24th International Conference on Machine Learning (ICML '07)*, pp. 209–216, June 2007.
- [21] E. P. Xing, M. I. Jordan, S. Russell, and A. Ng, "Distance metric learning with application to clustering with sideinformation," in *Advances in Neural Information Processing Systems*, vol. 15, pp. 505–512, 2002.
- [22] K. Q. Weinberger and L. K. Saul, "Distance metric learning for large margin nearest neighbor classification," *Journal of Machine Learning Research*, vol. 10, pp. 207–244, 2009.
- [23] H. Lian, W. Qiu, D. Yan, Z. Huang, and J. Guo, "Efficient Privacy-Preserving Protocol for," *Complexity*, vol. 2017, pp. 1–14, 2017.
- [24] N. Kumar, A. C. Berg, P. N. Belhumeur, and S. K. Nayar, "Attribute and simile classifiers for face verification," in *Proceedings of the 12th International Conference on Computer Vision (ICCV '09)*, pp. 365–372, IEEE, Kyoto, Japan, October 2009.
- [25] G. B. Huang, M. Ramesh, T. Berg, and E. Learned-Miller, *Labeled faces in the wild: a database for studying face recognition in unconstrained environments*, University of Massachusetts, Amherst, Mass, USA, 2007.
- [26] D. A. Freedman, "Statistical models: Theory and practice," *Statistical Models: Theory and Practice*, pp. 1–442, 2009.
- [27] V. Vapnik, *Statistical Learning Theory*, Wiley, New York, NY, USA, 1998.

- [28] M. Turk and A. P. Pentland, "Face recognition using eigenfaces," in *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, pp. 586–591, June 1991.
- [29] E. Nowak and F. Jurie, "Learning visual similarity measures for comparing never seen objects," in *Proceedings of the 2007 IEEE Computer Society Conference on Computer Vision and Pattern Recognition, CVPR'07, usa*, June 2007.
- [30] L. Wolf, T. Hassner, and Y. Taigman, "Descriptor based methods in the wild," in *Proceedings of the Faces in Real-Life Images Workshop in European Conference on Computer Visio*, 2008.
- [31] 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 *Proceedings of the 2009 IEEE Computer Society Conference on Computer Vision and Pattern Recognition Workshops, CVPR Workshops 2009*, pp. 2591–2598, usa, June 2009.
- [32] H. Li, G. Hua, Z. Lin, J. Brandt, and J. Yang, "Probabilistic elastic matching for pose variant face verification," in *Proceedings of the 26th IEEE Conference on Computer Vision and Pattern Recognition, CVPR 2013*, pp. 3499–3506, usa, June 2013.
- [33] F. Juefei-Xu, K. Luu, and M. Savvides, "Spartans: Single-sample periocular-based alignment-robust recognition technique applied to non-frontal scenarios," *IEEE Transactions on Image Processing*, vol. 24, no. 12, pp. 4780–4795, 2015.
- [34] M. Everingham, J. Sivic, and A. Zisserman, "'Hello! My name is... Buffy' - Automatic naming of characters in TV video," in *Proceedings of the 2006 17th British Machine Vision Conference, BMVC 2006*, pp. 899–908, gbr, September 2006.
- [35] J. Liu, S. Ji, and J. Ye, *SLEP: sparse learning with efficient projections*, Arizona State University, 2009, <http://www.public.asu.edu/jye02/Software/SLEP>.
- [36] C. Chang and C. Lin, "LIBSVM: a Library for support vector machines," *ACM Transactions on Intelligent Systems and Technology*, pp. 1–27, 2011, <http://www.csie.ntu.edu.tw/~cjlin/libsvm>.

