Graphical Representation for Heterogeneous Face Recognition

Chunlei Peng, Xinbo Gao, Nannan Wang, and Jie Li

Abstract—Heterogeneous face recognition (HFR) refers to matching non-photograph face images to face photos for identification. HFR plays an important role in both biometrics research and industry. In spite of promising progresses achieved in recent years, HFR is still a challenging problem due to the difficulty to represent two heterogeneous images in a homogeneous manner. Existing HFR methods either represent an image ignoring the spatial information, or rely on a transformation procedure which complicates the recognition task. Considering these problems, we propose a novel graphical representation based HFR method (G-HFR) in this paper. Markov networks are deployed to represent heterogeneous image patches separately, which take the spatial compatibility between neighboring image patches into consideration. A coupled representation similarity metric (CRSM) is designed to measure the similarity between obtained graphical representations. Extensive experiments conducted on two viewed sketch databases and a forensic sketch database show that the proposed method outperforms state-of-the-art methods.

Index Terms—Heterogeneous face recognition, Graphical representation, Forensic sketch.

I. INTRODUCTION

Face images captured through different sources, such as sketch artists and infrared imaging devices, are called in different modalities, i.e. heterogeneous faces. Matching these non-photograph face images to face photos, which is referred as heterogeneous face recognition (HFR), is now attracting growing attentions in both biometrics research and industry. For instance, there are circumstances where the photo of the suspect is not available and matching sketches to a large-scale database of mug shots is desired; Matching near infrared (NIR) images to visual (VIS) images is important for biometric security control to handle complicated illumination conditions.

Because of the textural discrepancies between heterogeneous face images, conventional homogeneous face recognition methods perform poorly by directly identifying the probe image (e.g. face sketch) from gallery images (e.g. face photos). Existing approaches can be generally grouped into three categories: synthesis based methods, common space projection based methods, and feature descriptor based methods. Synthesis based methods [1], [2], [3], [4], [5] first transform the heterogeneous face images into the same modality (e.g. photo). Once the synthesized faces are generated from non-photograph images or vice versa, conventional face recognition algorithms can be applied directly. However, the synthesis process is actually more difficult than recognition and the performance of these methods heavily depends on the fidelity of the synthesized images. Common space projection based methods [6], [7], [8], [9], [10], [11] attempt to project face images in different modalities into a common subspace where the discrepancy is minimized. Then heterogeneous face images can be matched directly in this common subspace. Yet the projection procedure always causes information loss which decreases the recognition performance. Feature descriptor based methods [12], [13], [14], [15], [16] first represent face images with local feature descriptors. These encoded descriptors can then be utilized for recognition. However, most existing methods of this category represent an image ignoring the special spatial structure of faces, which is crucial for face recognition in reality.

This paper proposes a novel graphical representation based HFR approach (G-HFR), which does not rely on any synthesis or projection procedure but takes spatial information into consideration. After face images are divided into overlapping patches, Markov networks are deployed to model the relationship between homogeneous image patches based on a representation dataset. The representation dataset consists of a number of heterogeneous face image pairs. Then graphical representations can be extracted from the Markov networks. Considering the spatial structure between heterogeneous face image patches, a coupled representation similarity metric (CRSM) is designed to measure the similarity between their graphical representations. Finally, calculated similarity scores between heterogeneous face images are applied for recognition.

The performance of the proposed G-HFR approach is thoroughly validated on two viewed sketch databases (the CUHK Face Sketch FERET Database (CUFSF) [16] and the CUHK Face Sketch Database (CUFS) [5]) and a forensic sketch database. Experimental results illustrate that the proposed approach achieves superior performance in comparison to state-of-the-art methods.

II. RELATED WORK

In this section, we briefly review representative HFR methods in aforementioned three categories: synthesis based methods, common space projection based methods, and feature descriptor based methods.

Synthesis based HFR methods began with an eigen-transformation algorithm [3] proposed by Tang and Wang. Later, Liu et al. [2] proposed a local linear embedding approach for patch-based face sketch synthesis. The sketch patches were synthesized independently and the spatial compatibility between neighboring patches was neglected. Chen et al. [17] proposed to learn the local linear mappings between NIR and VIS patches in a similar manner as [2]. Gao et al. [1] employed embedded hidden Markov model to represent
the non-linear relationship between sketches and photos and a selective ensemble strategy [18] was explored to synthesize a sketch. Wang and Tang [5] proposed a multi-scale Markov random field model for face sketch-photo synthesis, which took the spatial constraints between neighboring patches into consideration. Zhou et al. [19] proposed a Markov weight field model which was capable of synthesizing new patches that do not appear in the training set. Wang et al. [4] presented a transductive face sketch-photo synthesis method which incorporated the test image into the learning process.

In order to minimize the intra-modality difference, Lin and Tang [8] proposed a common discriminant feature extraction (CDE) approach to map heterogeneous features into a common feature space. The canonical correlation analysis (CCA) was applied to learn the correlation between NIR and VIS face images by Yi et al. [11]. Lei and Li [7] proposed a subspace learning framework for heterogeneous face matching, which was called coupled spectral regression (CSR). They later improved the CSR by learning the projections based on all samples from all modalities [20]. Sharma and Jacobs [10] used partial least squares (PLS) to linearly map images from different modalities to a common linear subspace. A cross modal metric learning (CMM) algorithm was proposed by Mignon and Jurie [9] to learn a discriminative latent space. Both the positive and negative constraints were considered in metric learning procedure. Kan et al. [6] proposed a multi-view discriminant analysis (MVDA) method to obtain a discriminant common space for recognition. The correlations from both inter-view and intra-view were exploited.

A number of feature descriptor based HFR approaches have shown promising performance. Klare et al. [15] proposed a local feature-based discriminant analysis (LFDA) framework through scale invariant feature transform (SIFT) feature [21] and multiscale local binary pattern (MLBP) feature [22]. A face descriptor based on coupled information-theoretic encoding was designed for matching face sketches with photos by Zhang et al. [16]. The coupled information-theoretic projection tree was introduced and was further extended to the randomized forest with different sampling patterns. Another face descriptor called local radon binary pattern (LRBP) was proposed in [12]. The face images were projected onto the radon space and encoded by local binary patterns (LBP). A histogram of averaged oriented gradients (HAOG) face descriptor was proposed to reduce the modality difference [13]. Lei et al. [23] proposed a discriminant image filter learning method benefitted from LBP like face representation for matching NIR to VIS face images. Alex et al. [24] proposed a local difference of Gaussian binary pattern (LDoGBP) for face recognition across modalities.

With great progresses achieved on viewed sketches, recently researches began to focus on matching forensic sketches to mug shots. Klare et al. [15] matched forensic sketches to mug shot photos with a populated gallery. Bhatt et al. [25] proposed a discriminative approach for matching forensic sketches to mug shots deploying multi-scale circular Weber’s local descriptor (MCWLD) and an evolutionary memetic optimization algorithm. Klare and Jain [14] represented heterogeneous face images through their nonlinear kernel similarities to a collection of prototype face images. Considering the fact that many law enforcement agencies employ facial composite software to create composite sketches, Han et al. [26] proposed a component based approach for matching composite sketches to mug shot photos.

III. Graphical Representation for Heterogeneous Face Recognition

In this section, we present a new approach for HFR. Without loss of generality and for ease of representation, we take face sketch-photo recognition as an example to describe the proposed method. Considering a representation dataset with \( M \) face sketch-photo pairs \( \{ (s^1, p^1), \ldots , (s^M, p^M) \} \), we first divide each face image into \( N \) overlapping patches. The probe sketch \( t \) and the gallery photos \( \{ g^1, \ldots , g^L \} \) are also divided into \( N \) overlapping patches respectively. Here \( L \) denotes the number of photos in the gallery. For a probe sketch patch \( y_i (i = 1, 2, \ldots , N) \), we can find \( K \) nearest sketch patches from the sketches in the representation dataset within the search region around the location of \( y_i \). The probe sketch patch \( y_i \) can then be regarded as a linear combination of the \( K \) nearest sketch patches \( \{ y_{i,1}, \ldots , y_{i,K} \} \) weighted by a column vector \( w_{y_i} = (w_{y_{i,1}}, \ldots , w_{y_{i,K}})^T \). The weight vector \( w_{y_i} \) is regarded as a representation of the probe sketch patch \( y_i \). For a gallery photo patch \( x^l_i \) from the \( l \)th gallery photo \( g^l \), where \( l = 1, 2, \ldots , L \), we can also find \( K \) nearest photo patches from the photos in the representation dataset and reconstruct the photo patch by a linear combination of these \( K \) nearest photo patches weighted by \( w_{x^l_i} \). The weight vector \( w_{x^l_i} \) is regarded as a representation of the gallery photo patch \( x^l_i \). The proposed approach is based on the observation that two heterogeneous face image patches corresponding to the same location from the same person tend to have similar representations, and the representations of two heterogeneous face image patches from different persons usually differ greatly.

The reconstruction weights can be simply generated through conventional subspace learning approaches such as principal component analysis (PCA) [27] and local linear embedding (LLE) [28]. However, these approaches neglect the spatial information which is essential for face recognition. In this paper, we propose to utilize Markov networks to represent heterogeneous face image patches separately, which take full advantage of the spatial compatibility between adjacent patches. Once graphical representations for probe sketch patches and gallery photo patches are obtained, a CRSM to measure the similarity between the probe sketch \( t \) and the gallery photo \( g^l \) is designed. Figure 1 gives an overview of the proposed method. The details are introduced as follows.

A. Graphical Representation

Inspired by the successful application of Markov networks on synthesis scenarios [5], [19], we jointly model all patches from a probe sketch or from a gallery photo on Markov networks. The joint probability of the probe sketch patches
and the weights is defined as
\[
p(w_{y_1}, \ldots, w_{y_N}, y_1, \ldots, y_N) = \prod_{i} \Phi(f(y_i), f(w_{y_j})) \prod_{(i,j) \in \Xi} \Psi(w_{y_i}, w_{y_j})
\]  
(1)

where \((i,j) \in \Xi\) denotes the \(i\)th probe sketch patch and the \(j\)th probe sketch patch are adjacent. \(f(y_i)\) means the feature extracted from the probe sketch patch \(y_i\), and \(f(w_{y_j})\) denotes the linear combination of features extracted from neighboring sketch patches in the representation dataset, \(i.e. f(w_{y_j}) = \sum_{k=1}^{K} w_{y_{i,k}} f(y_{i,k})\). \(\Phi(f(y_i), f(w_{y_j}))\) is the local evidence function, and \(\Psi(w_{y_i}, w_{y_j})\) is the neighboring compatibility function.

The local evidence function \(\Phi(f(y_i), f(w_{y_j}))\) is defined as
\[
\Phi(f(y_i), f(w_{y_j})) \propto \exp\{ -\|f(y_i) - \sum_{k=1}^{K} w_{y_{i,k}} f(y_{i,k})\|^2 / 2\delta_f^2 \}
\]  
(2)

The rationale behind the local evidence function is that \(\sum_{k=1}^{K} w_{y_{i,k}} f(y_{i,k})\) should be similar to \(f(y_i)\). Then the weight vector \(w_{y_i}\) is regarded as a representation of the probe sketch patch \(y_i\).

The neighboring compatibility function \(\Psi(w_{y_i}, w_{y_j})\) is defined as
\[
\Psi(w_{y_i}, w_{y_j}) \propto \exp\{ -\|\sum_{k=1}^{K} w_{y_{i,k}} o_{i,k} - \sum_{k=1}^{K} w_{y_{j,k}} o_{j,k}\|^2 / 2\delta_c^2 \}
\]  
(3)

where \(o_{i,k}\) represents the vector consisting of intensity values extracted from overlapping area (between the \(i\)th probe sketch patch and the \(j\)th probe sketch patch) in the \(k\)th nearest sketch patch of the \(i\)th probe sketch patch. The neighboring compatibility function is utilized to guarantee that neighboring patches have compatible overlaps.

Maximizing the joint probability function (1), we can obtain the optimal representations for the probe sketch. By substituting equations (2) and (3) into equation (1), maximizing the joint probability function (1) is equivalent to the minimization problem as follows
\[
\min_w \frac{1}{2\delta_f^2} \sum_{(i,j) \in \Xi} \| \sum_{k=1}^{K} w_{y_{i,k}} o_{i,k} - \sum_{k=1}^{K} w_{y_{j,k}} o_{j,k} \|^2 + \frac{1}{2\delta_c^2} \sum_{i=1}^{N} \| f(y_i) - \sum_{k=1}^{K} w_{y_{i,k}} f(y_{i,k}) \|^2
\]  
(4)

s.t.
\[
\sum_{k=1}^{K} w_{y_{i,k}} = 1, \ 0 \leq w_{y_{i,k}} \leq 1
\]  
\[
i = 1, 2, \ldots, N, \ k = 1, 2, \ldots, K
\]

where \(w\) is the concatenation of \(w_{y_i}(i = 1, 2, \ldots, N)\) in a long-vector form. Equation (4) can be further simplified as
\[
\min_w \alpha \sum_{(i,j) \in \Xi} \| O_i^2 w_{y_{i,k}} - O_j^2 w_{y_{j,k}} \|^2 + \sum_{i=1}^{N} \| f(y_i) - F_i w_{y_i} \|^2
\]  
(5)

where \(\alpha = \delta_f^2 / \delta_c^2\). \(F_i\) and \(O_i^2\) are two matrices, with the \(k\)th column being \(f(y_{i,k})\) and \(o_{i,k}\), respectively. Equation (5) can be rewritten as the following problem
\[
\min_w \ w^T Q w + w^T c + b
\]  
(6)

s.t.
\[
\sum_{k=1}^{K} w_{y_{i,k}} = 1, \ 0 \leq w_{y_{i,k}} \leq 1
\]  
\[
i = 1, 2, \ldots, N, \ k = 1, 2, \ldots, K
\]

where
\[
Q = \alpha \sum_{(i,j) \in \Xi} (O_i^2 - O_j^2)^T (O_i^2 - O_j^2) + \sum_{i=1}^{N} F_i^2 F_i
\]

\[
c = -2 \sum_{i=1}^{N} F_i^T f(y_i)
\]

\[
b = \sum_{i=1}^{N} f^T(y_i) f(y_i)
\]

The bias term \(b\) has no effect on the optimization problem and we can ignore it. The problem in equation (6) is optimized by
the cascade decomposition method [19] and then we obtain the weight matrix of the probe sketch $W_t = [w_{y,1}, \cdots, w_{y,N}]$. The weight matrix $W^{g_i} = [w^{g_i}_{1}, \cdots, w^{g_i}_{M}]$ of the $l$th gallery photo $g^l$ can be obtained in a similar way as aforementioned by jointly model all the gallery photo patches from $g^l$ and corresponding neighboring photo patches in the representation dataset.

To match the representation $w_{y_i}$ of a probe sketch patch to $w^{g_i}_{l}(l = 1, 2, \cdots, L)$, these weight vectors are reformulated as $M$-dimensional vectors. For the ease of denotations, these reformulated vectors are still represented as before. Each reformulated vector has at most $K$ nonzero values. For example, $w_{y,1}(z = 1, 2, \cdots, M)$ is nonzero only if the $z$th patch extracted from the $z$th sketch in the representation dataset is among the $K$ nearest neighbors of the probe sketch patch $y_i$.

### B. Coupled Representation Similarity Metric

In order to measure the similarity between two representations $W_t$ and $W^{g_i}$, we measure the similarity on each coupled patch pair respectively. Here “couple” means two column vectors from $W_t$ and $W^{g_i}$ have the same column order. L1 norm, L2 norm, and the cosine distance are three common used metrics to measure the similarity between two vectors. However, these metrics may be not the most suitable measure to calculate the similarity between graphical representations generated above due to their special characteristics, i.e., two representation vectors corresponding to the same position in different images share similar semantic meanings. Specially, $w_{y, z}$ and $w_{x^l, z}$ represent the weights of the sketch patch and photo patch from the $z$th ($z = 1, 2, \cdots, M$) sketch-photo pair in the representation dataset. Inspired by the rank-based similarity measure in [29], we propose a new similarity measure, namely coupled representation similarity metric (CRSM), to cater for this characteristic.

We compute the similarity score of the probe sketch patch $y_i$ and the gallery photo patch $x^l_i$ as the sum of the weights sharing the same nearest neighbors

\[
s(y_i, x^l_i) = 0.5 \sum_{z=1}^{M} n_z(w_{y, z} + w_{x^l, z})
\]

where

\[
n_z = \begin{cases} 
1, & w_{y, z} > 0 \text{ and } w_{x^l, z} > 0 \\
0, & \text{otherwise}
\end{cases}
\]

The proposed similarity measure ranges from 0 to 1.

The effect of the number of nearest neighbors $K$ on the similarity measurement is shown in Figure 2. We find that similarity map images corresponding to heterogeneous faces of the same person tend to have more bright areas than those of different persons have. The average of the similarity scores on all patch positions can be regarded as the final similarity score between the probe sketch and the gallery photo. In Figure 2, the numbers below the similarity map images are similarity scores obtained.

### IV. Experiments

In this section, we evaluated the performance of the proposed approach on the task of face sketch-photo recognition. We first evaluated the effectiveness of the proposed graphical representation and the effectiveness of CRSM separately. Then we investigated the effect of different parameters and number of features on the recognition performance. Finally we validated that our approach achieved superior performance compared with state-of-the-art methods on three databases: the CUHK face sketch FERET database (CUFSF) [16], the CUHK face sketch database (CUFS) [5], and a forensic sketch database.

#### Databases

CUFSF includes 1194 sketch-photo pairs with photos collected from the FERET database [30]. There are lighting variations in the photos and shape exaggerations in the sketches of this database.

CUFS contains 606 sketch-photo pairs, with photos collected from the CUHK student database, the AR database [31] and the XM2VTS database [32]. All the photos are frontal, with a neutral expression and normal lighting conditions, and the sketches are drawn without exaggeration.

The forensic sketch database contains 168 real world forensic sketches and corresponding mug shot photos. The forensic sketches are drawn by sketch artists with the descriptions of eyewitnesses or victims. This database originates from a collection of images from the forensic sketch artist Lois Gibson [33], the forensic sketch artist Karen Taylor [34], and other internet sources. Some example images from these databases are shown in Figure 3.

On the CUFSF database, 250 persons are randomly selected as the representation dataset, and 250 persons are randomly selected as the set for training classifiers (namely training set). The remaining 694 persons form the testing set. On the CUFS database, 153 persons are in the representation dataset, and 153 persons are in the training set. The remaining 300 persons form the testing set. On the forensic sketch database, the CUHK AR database including 123 sketch-photo pairs is chosen as
the representation dataset. We follow the similar partition in [14] and 112 persons from the forensic sketch database are randomly selected as the training set. The remaining 56 persons are used for testing. We randomly selected 10000 face photo images from the labeled faces in the wild-a (LFW-a) database [29] to increase the scale of the gallery, which mimic the real-world face retrieval scenarios in law enforcement agencies. LFW-a is a database of labeled face images in the LFW dataset [35]. Example faces are shown in Figure 3. Note that the accuracies reported in this paper are statistical results over 10 random partitions of the training set and testing set.

A. Experimental Settings

The parameters appeared in this paper are set as follows. Each face image is cropped to 100\times125. The image patch size is 10, and the overlapping size is 5, i.e. there are 456 patches per image. The neighborhood search region is 16\times16. In the Markov networks, we do not set \beta_\Phi or \beta_\Psi directly, but instead \alpha is set to 0.025, where \alpha = \beta_\Phi/\beta_\Psi. To determine other experimental settings, we conducted adjustment experiments on the CUFSF database. Once these experimental settings are determined, they are kept constant in following experiments.

To illustrate the effectiveness of graphical representations, we replace the Markov networks with the local linear embedding [28] which ignores the spatial information. Speeded up robust features (SURF) [36] are utilized as the feature descriptor and the number of the nearest neighbors K is set to 15. As shown in the left top subfigure of Figure 4, the spatial information is essential for HFR.

To illustrate the effectiveness of the proposed similarity metric (CRSM), we compare it with L1 norm, L2 norm and the cosine distance. SURF is utilized as the feature descriptor and K is set to 15. As shown in the right top subfigure of Figure 4, the L2 norm performs poorly on proposed graphical representations. The proposed similarity measure is more effective than L1 norm and cosine distance.

We test the effect of the number of nearest neighbors K with SURF as the feature descriptor. K is set to 15, 20, 25, 30, 35 and 40 respectively. As shown in the left bottom subfigure of Figure 4, the recognition accuracy varies with different K values.

In our experiments, we find that fusion of different similarity metrics corresponding to different K values would further improve the performance. We explore a linear one-class SVM to fuse the similarity scores obtained by different K values. We follow the fusion strategy in [16] and select all the intrapersonal pairs and the same number of interpersonal pairs with largest similarity scores to train the one-class SVM. As shown in the right bottom subfigure of Figure 4, the increase of the number of similarity metrics does improve the recognition accuracy. The rationale behind this is that complementary information exists among different similarity metrics. Combining 6 similarity metrics increases the accuracy from 92.22\% to 94.24\%.

We also investigate the effect of the fusion of different features on the recognition performance. Because the proposed method represents the heterogeneous face images in each modality separately, common features used in homogeneous face recognition are sufficient for the task. In this paper, SURF [36], SIFT [21], and histograms of oriented gradients (HOG) [37] are deployed to represent an image patch respectively.

To illustrate the necessity of spatial information, right top subfigure shows the comparison of the CRSM with traditional similarity metrics; left bottom subfigure shows the accuracies of different numbers of the nearest neighbors K; right bottom subfigure shows the accuracies by fusion of similarity metrics. All the four experiments are conducted on the CUFSF database using the SURF feature.

B. Experiments on the CUFSF Database

We compare the proposed G-HFR method with state-of-the-art approaches on the CUFSF database as shown in Table I. For
the transductive synthesis method (TFSPS) [4], query sketches are transformed into synthesized photos, and random sampling LDA (RS-LDA) [38] is used to match the synthesized photos to gallery photos. Because the photos and sketches in CUFSF involve lighting variations and shape exaggerations, the synthesized photos have artifacts such as distortions. These artifacts degrade the recognition performance. For the common space projection based approaches PLS [10] and MvDA [6], a discriminant common space for two modalities is learnt. Though these two approaches have a strong generality and can be applied to various heterogeneous scenarios, they perform poorly on CUFSF as shown in the table. For feature descriptor based methods LRBP [12] and LDoGBP [24], feature descriptors which are invariant to different modalities are designed and used for recognition. These two approaches achieve good performance with accuracies of 91.12% and 91.04% respectively. However, these features ignore the spatial structure of faces. Our proposed method achieves a first match rate of 96.04% and a tenth match rate of 99.86%. Zhang et al. [16] achieved 98.70% verification rates (VR) at 0.1% false acceptance rate (FAR) in comparison to 99.14% VR at 0.1% FAR of our proposed G-HFR method.

C. Experiments on the CUFS Database

G-HFR achieves nearly perfect performance on the CUFS database, which is a standard benchmark for face sketch-photo recognition. As shown in Table II, recognition accuracies on this benchmark are saturated. Many state-of-the-art methods achieve accuracies higher than 99%. The HAOG [13] method emphasizes coarse texture of facial components in feature extraction stage and achieves 100% accuracy on this benchmark. Our method achieves a first match rate of 99.90%.

D. Experiments on the Forensic Sketch Database

Matching forensic sketches to mug shots is much more difficult than matching aforementioned viewed sketches, because forensic sketches are drawn based on the eyewitness’s descriptions. This can be easily affected by various eyewitnesses’ face perceptions and sketch artists’ perceptual experiences when drawing the forensic sketches. It is even harder when the eyewitness’s description contains verbal overshadowing and memory distorting properties.

Table III compares the recognition performance of the proposed G-HFR method and other state-of-the-art methods. The LFDA [15] method fuses SIFT and MBLP with a discriminative model. They improve the accuracy by applying an ensemble of discriminant classifiers. With a collection of 159 forensic sketches, they perform matching against a populated gallery with 10159 mug shots. They achieve a rank-50 accuracy of 13.4%, which is the weighted average of an accuracy of 32.65% on 49 good sketches and 8.16% on 110 poor sketches. The MCWLD [25] method utilizes 140 semi-forensic sketches for training and 190 forensic sketches are taken as the probe images. 599 face photos plus 6324 photos form the gallery. A rank-50 accuracy of 28.52% is achieved by this method. For the prototype random subspaces (P-RS) [14] method, three different image filters and two different local feature descriptors are applied to the probe and gallery images. A set of prototypes representing both the probe and gallery modalities are used for training and a random subspace framework is employed to boost the performance. They utilized 106 subjects for training and 53 subjects plus 10000 mug shots for testing and achieved a rank-50 accuracy of 20.80%. The component based approach [26] uses 123 composite sketches as the probe set and 10000 mug shots are added to the gallery. They achieve a rank-50 accuracy of 52.03%. Note that the composite sketches are generated with each component approximated by the most similar component available in the composite software’s database. Our method achieves a rank-50 accuracy of 57.86%, which outperforms existing state-of-the-art methods. The cumulative match scores of the proposed method and the methods [25], [26], [14] are shown in Figure 6. Due to the small scale of available forensic sketch database, there are not enough sketches for training a strong model. It is reasonable to believe that the recognition performance can be further improved with more forensic sketches available.

V. Conclusions

A graphical representation based heterogeneous face recognition method (G-HFR) is proposed in this paper. G-HFR deploys Markov networks to represent heterogeneous face

**Table I**

| Method       | Accuracy |
|--------------|----------|
| TFSPS [4]    | 72.62%   |
| MvDA [6]     | 55.50%   |
| LDoGBP [24]  | 91.04%   |
| PLS [10]     | 51%      |
| LRBP [12]    | 91.12%   |
| G-HFR        | 96.04%   |

**Table II**

| Method       | Accuracy |
|--------------|----------|
| Eigensketch [3] | 90%        |
| MRF [5]      | 96.30%   |
| TFSPS [4]    | 97.70%   |
| PLS [10]     | 93.60%   |
| LFDA [15]    | 99.47%   |
| CITE [16]    | 99.87%   |
| LRBP [12]    | 99.51%   |
| HAOG [13]    | 100%     |
| LDoGBP [24]  | 96.53%   |
| G-HFR        | 99.90%   |

**Table III**

| Method       | Accuracy |
|--------------|----------|
| Eigensketch [3] | 90%        |
| MRF [5]      | 96.30%   |
| TFSPS [4]    | 97.70%   |
| PLS [10]     | 93.60%   |
| LFDA [15]    | 99.47%   |
| CITE [16]    | 99.87%   |
| LRBP [12]    | 99.51%   |
| HAOG [13]    | 100%     |
| LDoGBP [24]  | 96.53%   |
| G-HFR        | 99.90%   |

Fig. 5. Experiments on using different features and the fusion of them on the CUFSF database.
images with the spatial information taken into consideration. Considering the coupled spatial property between heterogeneous face image patches, we propose a coupled representation similarity metric. Experiments are conducted to illustrate the effect of the proposed graphical representation and similarity metric in comparison to common used representations and similarity metrics. Compared with state-of-the-art methods on three heterogeneous face databases, G-HFR achieves superior performance in terms of face recognition accuracy. In the future, the effect of more types of features would be investigated to further improve the recognition performance. Furthermore, we would further evaluate the performance of the proposed G-HFR method on more heterogeneous face recognition scenarios.

REFERENCES

[1] X. Gao, J. Zhong, J. Li, and C. Tian, “Face sketch synthesis algorithm based on E-HMM and selective ensemble,” IEEE TCSVT, vol. 18, no. 4, pp. 487–496, 2008.
[2] Q. Liu, X. Tang, H. Jin, H. Lu, and S. Ma, “A nonlinear approach for face sketch synthesis and recognition,” in CVPR, 2005.
[3] X. Tang and X. Wang, “Face sketch synthesis and recognition,” in ICCV, 2003.
[4] N. Wang, D. Tao, X. Gao, X. Li, and J. Li, “Transductive face sketch-photo synthesis,” IEEE TNNLS, vol. 24, no. 9, pp. 1364–1376, 2013.
[5] X. Wang and X. Tang, “Face photo-sketch synthesis and recognition,” IEEE TPAMI, vol. 31, no. 11, pp. 1955–1967, 2009.
[6] M. Kan, S. Shan, H. Zhang, S. Lao, and X. Chen, “Multi-view discriminant analysis,” in ECCV, 2012.
[7] Z. Lei and S. Li, “Coupled spectral regression for matching heterogeneous faces,” in CVPR, 2009.
[8] D. Lin and X. Tang, “Inter-modality face recognition,” in ECCV, 2006.
[9] A. Mignon and F. Jurie, “CMML: a new metric learning approach for cross modal matching,” in ACCV, 2012.
[10] A. Sharma and D. Jacobs, “Bypass synthesis: PLS for face recognition with pose, low-resolution and sketch,” in CVPR, 2011.
[11] D. Yi, R. Liu, R. Chu, Z. Lei, and S. Li, “Face matching between near infrared and visible light images,” in ICB, 2007.
[12] H. Galkoogahi and T. Sim, “Face sketch recognition by local radon binary pattern,” in ICP, 2012.
[13] ——, “Inter-modality face sketch recognition,” in ICME, 2012.