Cross-Modality Face Recognition via Heterogeneous Joint Bayesian

Hailin Shi, Xiaobo Wang, Dong Yi, Zhen Lei, *Senior Member, IEEE*, Xiangyu Zhu, and Stan Z. Li, *Fellow, IEEE*

***Abstract*—In many face recognition applications, the modalities of face images between the gallery and probe sets are different, which is known as heterogeneous face recognition. How to reduce the feature gap between images from different modalities is a critical issue to develop a highly accurate face recognition algorithm. Recently, joint Bayesian (JB) has demonstrated superior performance on general face recognition compared to traditional discriminant analysis methods like subspace learning. However, the original JB treats the two input samples equally and does not take into account the modality difference between them and may be suboptimal to address the heterogeneous face recognition problem. In this work, we extend the original JB by modeling the gallery and probe images using two different Gaussian distributions to propose a heterogeneous joint Bayesian (HJB) formulation for cross-modality face recognition. The proposed HJB explicitly models the modality difference of image pairs and, therefore, is able to better discriminate the same/different face pairs accurately. Extensive experiments conducted in the case of visible–near-infrared and ID photo versus spot face recognition problems show the superiority of the HJB over previous methods.**

***Index Terms*—Cross modality, heterogeneous face recognition, joint Bayesian (JB).**

I. Introduction

HETEROGENEOUS face recognition is a common issue in many face recognition applications, where the gallery and the probe face images come from different modalities. For example, in the application of access control, the gallery is usually controlled visible (VIS) photo, while the probe sometimes prefers to be near-infrared (NIR) image, which is robust to illumination variations [1]. This is a cross-modality face recognition problem between VIS and NIR face images. In the remote face verification, which is increased in recent years, the gallery is usually the ID photo captured in the constrained condition, while the probe is the face image captured by a cellphone or a webcam in a more arbitrary environment, which contains more variations of lighting, pose, expression, accessory, etc.

Up to now, many approaches have been proposed to address the heterogeneous face recognition problem. One category is to extract modality-invariant features to reduce the feature gap between different modalities so that the face images from different modalities can be well matched. Liao *et al.* [2] use difference of Gaussian to obtain the normalized appearances from different modalities and uses multiscale block local binary pattern to extract discriminative features. Zhang *et al.* [3] propose a face descriptor based on coupled information-theoretic encoding to capture the local face structure of photo and sketch face images. Liu *et al.* [4] derive light-source-invariant features to extract invariant parts between the different modalities. Lei *et al.* [5] extend the discriminative local features in a coupled way to reduce the difference between features of heterogeneous face images.

Another sort of methods focus on coupled metric or classifier learning. Lin and Tang [6] propose to learn two transforms simultaneously and transform the inputs from different modalities to a common features space. LSR-LDA [7] copes with the irregular distribution of heterogeneous face data to improve the conventional LDA. Lei *et al.* [8], [9] propose to learn two coupled projections to map the face images from different modalities to a common subspace, in which good discrimination can be gained. Klare *et al.* [10], [11] try to learn multiple projections for forensic sketch–photo matching. MCA [12] uses a learned generative model to infer the mutual components of different modalities.

Besides, researchers also propose to deal with heterogeneous face recognition in an analysis-by-synthesis way. The face analogy [13] performs heterogeneous face matching by transforming face images from one modality to another. Xu *et al.* [14] propose to reconstruct face images from each other's modality by using a learnt ℓ-0 minimization-based dictionary. Other methods [15], [16] apply the depth information and local binary pattern to accomplish the recognition task.

In recent years, deep learning methods have achieved great success in many computer vision tasks including face recognition. Certain deep convolutional neural network (CNN) models have been successively applied for general face recognition, such as DeepID2 [17], VGG Face [18], etc. There are also some pioneering works to address cross-modality face recognition by using deep learning methods. Yi *et al.* [19] use the Gabor feature and RBM to learn shared representation in order to reduce the heterogeneity in the encoder layer. Ensemble ELM [20] and MTC-ELM [21] employ the extreme learning machine for the feature learning of cross-modality face images. TRIVET [22] pretrains a deep CNN on a large dataset of general human face and finetunes it on the heterogenous face dataset.

Manuscript received September 22, 2016; revised November 24, 2016; accepted November 29, 2016. Date of publication December 8, 2016; date of current version January 5, 2017. This work was supported by the National Key Research and Development Plan under Grant 2016YFC0801002, the Chinese National Natural Science Foundation Projects 61473291, 61572501, 61502491, and 61572536, and AuthenMetric R&D Funds. The associate editor coordinating the review of this manuscript and approving it for publication was Prof. Jing-Ming Guo.

H. Shi, X. Wang, X. Zhu, and S. Z. Li are with the Institute of Automation, Chinese Academy of Sciences, Beijing 100190, China (e-mail: ncushl@gmail.com; xiaobo.wang@nlpr.ia.ac.cn; xiangyu.zhu@nlpr.ia.ac.cn; szli@nlpr.ia.ac.cn).

D. Yi is with Alibaba Group, Hangzhou 311121, China (e-mail: yidong.yd@alibabainc.com).

Z. Lei is with the National Laboratory of Pattern Recognition, Institute of Automation, Chinese Academy of Sciences, Beijing 100190, China (e-mail: zlei@nlpr.ia.ac.cn).

Color versions of one or more of the figures in this are available online at http://ieeexplore.ieee.org.

Digital Object Identifier 10.1109/LSP.2016.2637400

Recently, the joint Bayesian (JB) [23] method is proposed to model the intra- and interface pairs effectively for general face recognition. As a metric learning method, the JB method achieves superior accuracies of recognition with both the traditional features [23] and the deep learning features [17]. However, the JB method does not take into account the heterogeneity in cross-modality face recognition. Inspired by the effectiveness of the JB method, we extend it to the range of heterogeneous face recognition.

To this end, in this paper, we reformulate the JB method in an asymmetric form, namely heterogeneous joint Bayesian (HJB), in which the heterogeneity is taken into account for learning a more effective metric across different modalities. The HJB considers the two inputs as the samplings from two different Gaussian distributions and optimize the asymmetric metric with respect to the log-likelihood ratio across modalities. In this way, the HJB surpasses the baseline JB and achieves the state-of-the-art performance for the heterogeneous face recognition. Extensive experiments on NIR-VIS and ID photo versus spot faces validate the superiority of the HJB.[1]

The remainder of this paper is organized as follows. In Section II, we revisit the ordinary JB method. In Section III, we introduce the novel formulation of the HJB and its solution using the expectation–maximization (EM) method. In Section IV, we conduct the comparison of the HJB with previous methods on several benchmarks of NIR versus VIS and ID versus spot face recognition. We conclude the paper in Section V.

II. Revisit of JB

Let x be the representation of human face image. x is supposed to be comprised by two independent random variables μ and ϵ, i.e., $x = \mu + \epsilon$. The variables μ and ϵ represent the identity and the intraclass variations (e.g., pose, expression, illumination, etc.). As described in the previous works [24], [25], μ and ϵ can be regarded as two independent zero-mean Gaussian variables, i.e., $\mu \sim \mathcal{N}(0, S_\mu)$ and $\epsilon \sim \mathcal{N}(0, S_\epsilon)$. As the sum of μ and ϵ, x follows the Gaussian distribution $\mathcal{N}(0, S_\mu + S_\epsilon)$ as well. Considering two inputs x_1 and x_2, their joint distribution is also Gaussian. Denote by H_I the hypothesis the two inputs belong to the same subject, and by H_E the hypothesis of different subjects. One can write the covariance matrix of the intraclass joint distribution $P(x_1, x_2|H_I)$ as

$$\Sigma_I = \begin{bmatrix} S_\mu + S_\epsilon & S_\mu \\ S_\mu & S_\mu + S_\epsilon \end{bmatrix} \tag{1}$$

and the counterpart of the interclass joint distribution $P(x_1, x_2|H_E)$ as

$$\Sigma_E = \begin{bmatrix} S_\mu + S_\epsilon & 0 \\ 0 & S_\mu + S_\epsilon \end{bmatrix}. \tag{2}$$

The assumption behind this neat formulation is that the identity μ and the intraclass variations ϵ are independent. To measure the similarity between x_1 and x_2, the log-likelihood ratio is computed by

$$r(x_1, x_2) = \log \frac{P(x_1, x_2|H_I)}{P(x_1, x_2|H_E)} = x_1^T A x_1 + x_2^T A x_2 - 2 x_1^T G x_2. \tag{3}$$

One can refer to the original proposal [23] for the calculation details of the matrices A and G.

III. Heterogeneous Joint Bayesian

In this section, we introduce the asymmetric formulation of the HJB and the solution via the EM algorithm.

A. Asymmetric Model

By breaking the x_1–x_2 symmetry in the original JB, we introduce the gallery x and the probe y as two different random variables, and their decompositions as $x = \mu_x + \epsilon_x$ and $y = \mu_y + \epsilon_y$, respectively. The variables μ_x and μ_y are the identity variations, and ϵ_x and ϵ_y are the intraclass variations, all of which follow the zero-mean Gaussians, i.e., $\mu_x \sim \mathcal{N}(0, S_{xx})$, $\mu_y \sim \mathcal{N}(0, S_{yy})$, $\epsilon_y \sim \mathcal{N}(0, T_{xx})$, and $\epsilon_x \sim \mathcal{N}(0, T_{yy})$. Here, S_{xx}, S_{yy}, T_{xx}, and T_{yy} are the corresponding covariances, respectively. To reveal the connection between the gallery and probe, we introduce the covariance of the cross-modality identity variations between x and y as

$$S_{xy} = \mathrm{cov}(\mu_x, \mu_y) \tag{4}$$

$$S_{yx} = \mathrm{cov}(\mu_y, \mu_x) \tag{5}$$

which are mutual transposes $S_{xy} = S_{yx}^T$. The $\mathrm{cov}(\cdot,\cdot)$ denotes the covariance. Consequently, the covariance matrix of the intraclass joint distribution $P(x, y|H_I)$ is written as

$$\Sigma_I = \begin{bmatrix} S_{xx} + T_{xx} & S_{xy} \\ S_{yx} & S_{yy} + T_{yy} \end{bmatrix} \tag{6}$$

and the counterpart of the interclass joint distribution $P(x, y|H_E)$ is written as

$$\Sigma_E = \begin{bmatrix} S_{xx} + T_{xx} & 0 \\ 0 & S_{yy} + T_{yy} \end{bmatrix}. \tag{7}$$

With these covariance matrices, we revise the cross-modality log-likelihood ratio of x and y as

$$r(x, y) = \log \frac{P(x, y|H_I)}{P(x, y|H_E)} = x^T A x + y^T B y - 2 x^T G y \tag{8}$$

where

$$A = (S_{xx} + T_{xx})^{-1} - E \tag{9}$$

$$B = (S_{yy} + T_{yy})^{-1} - F \tag{10}$$

$$\begin{bmatrix} E & G \\ G^T & F \end{bmatrix} = \begin{bmatrix} S_{xx} + T_{xx} & S_{xy} \\ S_{yx} & S_{yy} + T_{yy} \end{bmatrix}^{-1}. \tag{11}$$

B. Solution

Based on the learning process in [23], we develop the EM-fashion algorithm to estimate the covariances S_{xx}, S_{yy}, T_{xx}, T_{yy}, and S_{xy} for each modality separately.

1) E-Step: We introduce two latent variables $\mathbf{h}_x$ and $\mathbf{h}_y$, composed of $\mathbf{h}_x = [\mu_x, \epsilon_{x,1}, \ldots, \epsilon_{x,n_g}]^T$ and $\mathbf{h}_y = [\mu_y, \epsilon_{y,1}, \ldots, \epsilon_{y,n_p}]^T$, corresponding to the galleries $\mathbf{x} = [x_1, \ldots, x_{n_g}]^T$ and probes $\mathbf{y} = [y_1, \ldots, y_{n_p}]^T$, respectively, of each subject.

Considering the decomposition of identity variations and intraclass variations, the galleries and the probes can be represented

[1]The source code of HJB will be released at http://www.cbsr.ia.ac.cn/users/hailinshi/

by the latent variables as $\mathbf{x} = \mathbf{P}_x \mathbf{h}_x$ and $\mathbf{y} = \mathbf{P}_y \mathbf{h}_y$, where $\mathbf{P}_x$ and $\mathbf{P}_y$ are the matrices with the form of

$$\begin{bmatrix} \mathbf{I} & \mathbf{I} & 0 & \dots & 0 \\ \mathbf{I} & 0 & \mathbf{I} & \dots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ \mathbf{I} & 0 & 0 & \dots & \mathbf{I} \end{bmatrix}.$$

$\mathbf{I}$ is the identity matrix. Obviously, the latent variables also follow the Gaussian distributions. Based on the algorithm in [23], we compute the expectation of latent variables in each modality by

$$E(\mathbf{h}_x|\mathbf{x}) = \Sigma_{\mathbf{h}_x} \mathbf{P}_x^T \Sigma_x^{-1} \mathbf{x}$$
$$= \begin{bmatrix} S_{xx} & 0 & \dots & 0 \\ 0 & T_{xx} & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & T_{xx} \end{bmatrix} \begin{bmatrix} \mathbf{I} & \mathbf{I} & \dots & \mathbf{I} \\ \mathbf{I} & 0 & \dots & 0 \\ 0 & \mathbf{I} & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & \mathbf{I} \end{bmatrix}$$
$$\begin{bmatrix} S_{xx}+T_{xx} & S_{xx} & \dots & S_{xx} \\ S_{xx} & S_{xx}+T_{xx} & \dots & S_{xx} \\ \vdots & \vdots & \ddots & \vdots \\ S_{xx} & S_{xx} & \dots & S_{xx}+T_{xx} \end{bmatrix}^{-1} \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_{n_g} \end{bmatrix} \tag{12}$$

$$E(\mathbf{h}_y|\mathbf{y}) = \Sigma_{\mathbf{h}_y} \mathbf{P}_y^T \Sigma_y^{-1} \mathbf{y}$$
$$= \begin{bmatrix} S_{yy} & 0 & \dots & 0 \\ 0 & T_{yy} & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & T_{yy} \end{bmatrix} \begin{bmatrix} \mathbf{I} & \mathbf{I} & \dots & \mathbf{I} \\ \mathbf{I} & 0 & \dots & 0 \\ 0 & \mathbf{I} & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & \mathbf{I} \end{bmatrix}$$
$$\begin{bmatrix} S_{yy}+T_{yy} & S_{yy} & \dots & S_{yy} \\ S_{yy} & S_{yy}+T_{yy} & \dots & S_{yy} \\ \vdots & \vdots & \ddots & \vdots \\ S_{yy} & S_{yy} & \dots & S_{yy}+T_{yy} \end{bmatrix}^{-1} \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_{n_p} \end{bmatrix}. \tag{13}$$

On the right-hand side of (12) and (13), $\Sigma_{\mathbf{h}_x}$ and $\Sigma_{\mathbf{h}_y}$ are the covariances of the latent variables $\mathbf{h}_x$ and $\mathbf{h}_y$. Σ_x is the covariance matrix of the joint distribution of the intraclass set $\mathbf{x}$. Σ_y is the counterpart for $\mathbf{y}$. To start the E-step, the parameters S_{xx}, S_{yy}, T_{xx}, and T_{yy} are initialized by the inter- and intraclass covariances of the training set.

2) M-Step: Once the latent variables $\mathbf{h}_x$ and $\mathbf{h}_y$ are estimated in the E-step, we compute the covariances of μ_x, μ_y, ϵ_x, and ϵ_y and use them to update the parameters, i.e., $S_{xx} = \mathrm{cov}(\mu_x, \mu_x)$, $S_{yy} = \mathrm{cov}(\mu_y, \mu_y)$, $T_{xx} = \mathrm{cov}(\epsilon_x, \epsilon_x)$, $T_{yy} = \mathrm{cov}(\epsilon_y, \epsilon_y)$, and $S_{xy} = \mathrm{cov}(\mu_x, \mu_y)$.

We train the model with the EM algorithm for a few iterations when the algorithm converges (generally in two iterations). Then, we use formulas (9)–(11) to compute the model components A, B, and G, and the log-likelihood ratio [see (8)] for testing.

IV. EXPERIMENTS

We examine the performance of the HJB compared with the previous methods including LCKS-CSR [26], MTC-ELM [21], NIR-VIS Reconstruction + UDP (DLBP) [14], and other state-of-the art methods on three databases, i.e., CASIA-HFB [26], CASIA NIR-VIS 2.0 [27], and a private database consisting of ID photo and spot face images.

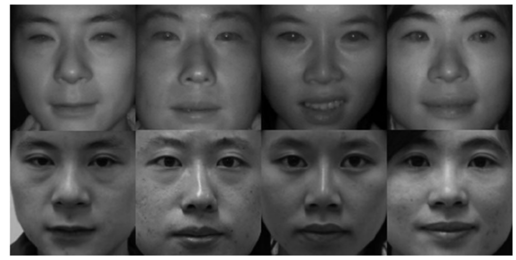

Fig. 1. CASIA-HFB database. (Top) NIR images. (Bottom) VIS images. Each column belongs to an identity.

TABLE I
PERFORMANCE COMPARISON WITH THE PREVIOUS METHODS ON CASIA-HFB

	Rank-1 Accuracy	VR @FAR = 10%	VR @FAR = 1%	VR @FAR = 0.1%
LDA [24]	72.43%	48.75%	26.55%	14.04%
CDFE [6]	16.10%	40.05%	12.75%	3.41%
LDA + CCA [28]	72.65%	42.90%	25.76%	13.79%
LCSR [8]	81.12%	71.28%	51.07%	33.98%
LCKS-CDA [26]	73.18%	54.11%	31.21%	16.61%
LCKS-CSR [26]	81.43%	75.18%	54.81%	**35.69**%
JB	82.30%	73.75%	50.31%	18.19%
HJB	**85.49**%	**80.82**%	**59.30**%	33.65%

A. *Experiments on CASIA-HFB*

CASIA HFB contains 300 subjects, with around five NIR images and five VIS images per subject. In this part, we follow the same protocol as in [26]. We use the images from the first 150 persons to form the training set and the left images to form the testing set. All the images are cropped into 32×32 gray images according to automatically detected eye locations. Some cropped examples are shown in Fig. 1. The pixel intensity is directly used as input. In testing, the VIS images are used as the gallery set and the NIR ones are used as the probe set.

We compare the HJB method with previous methods including the traditional homogenous method like LDA [24] and heterogeneous face recognition methods including LCSR [8], LDA + CCA [28], CDFE [6], LCKS-CDA [26], and LCKS-CSR [26] methods. The face recognition is evaluated in terms of rank-1 accuracy and receiver operating characteristic (ROC) performance.

Table I shows the rank-1 accuracy and verification rate (VR) at different false accept rates (FARs). From the results, one can see the following.

1) For the homogeneous face recognition methods, the original JB achieves better performance than LDA, indicating that the JB has good ability as a baseline for the heterogeneous face recognition. This leads to the basic motivation that we develop the HJB to exploit its advantage.
2) Comparing the HJB with the JB, one can see that the HJB achieves significantly better performance than the JB, especially at the low FAR. The HJB enhances the JB about 7–15% in VR with different FARs. It validates that the HJB does improve the heterogeneous face recognition performance by taking into account the modality difference in the learning process.

TABLE II
PERFORMANCE COMPARISON WITH THE STATE OF THE ART ON CASIA NIR-VIS 2.0

	Rank-1 Accuracy	VR@FAR = 0.1%
Cognitec [29]	58.56 ± 1.19%	N/A
CDFL [30]	71.5 ± 1.4%	55.1%
DSIFT + LDA [29]	73.28 ± 1.10%	N/A
Gabor + RBM [19]	86.16 ± 0.98%	81.29 ± 1.82%
NIR-VIS Reconstruction + UDP (DLBP) [14]	78.46 ± 1.67%	85.80%
MTC-ELM [21]	89.1%	N/A
TRIVET [22]	**95.7** ± 0.5%	**91.0** ± 1.3%
Gabor + JB	89.45 ± 0.79%	83.28 ± 1.03%
Gabor + HJB	**91.65** ± 0.89%	**89.91** ± 0.97%

3) The HJB outperforms previous heterogeneous face recognition methods in most cases. It improves about 4% over the previously best method (LCKS-CSR), validating that the HJB is an effective method to address the heterogeneous face recognition problem.

B. Experiments on CASIA NIR-VIS 2.0

The CASIA NIR-VIS 2.0 [27] is the largest and most popular database for the NIR–VIS face recognition task. It contains 725 subjects, each of which has 1–22 VIS and 5–50 NIR face images. Under the View2 protocol, the evaluation is performed via the tenfold process, In each fold, there are 357 subjects for training, and the remaining 358 subjects for test. We compare the proposed HJB with the previous methods, which give the state-of-the-art performances on CASIA NIR-VIS 2.0. Considering the good performance in [19], we use the local Gabor features as the inputs of the proposed HJB. We also compare the performance of our HJB with the baselines, i.e., the original JB.

Table II shows the performance of different methods on the NIR-VIS 2.0 database. The result reveals the following.

1) The general face recognition method proposed by Cognitec gives poor performances compared with the heterogeneous methods. It is critical to take into account the difference between modalities for heterogeneous face recognition.
2) Compared with Gabor + RBM [19], Gabor + HJB gains significantly better performance. It improves the RBM method by about 5% in both rank-1 and verification performance, validating the superiority of the HJB compared to RBM.
3) The HJB outperforms the previous state-of-the-art methods except the method TRIVET, which trains a deep CNN on a large outside dataset. Without any help of CNN, our HJB shows its effectiveness for the heterogeneous face recognition.

C. Experiments on ID Versus Spot Recognition

To further evaluate the HJB, we collect an ID versus spot face dataset, with 10 000 identities in it. Each identity has an ID photo and a spot photo (see Fig. 2). The ID photos and the spot photos are captured under different conditions (i.e., the lightening, background, pose, etc.). This is a very challenging dataset due to the significant difference between the modalities, and the large variations in the spot set.

Because each subject has only one ID photo (gallery) and one spot photo (probe), we are not able to estimate the intraclass covariances ϵ_x and ϵ_y. Instead, we suppose ϵ_x and ϵ_y are not random but determined entities. Therefore, the components T_{xx} and T_{yy} vanish in the relevant computation, keeping the rest part unchanged.

Fig. 2. (Top) ID photos. (Bottom) Spot photos. Each column belongs to an identity.

TABLE III
PERFORMANCES ON ID VERSUS SPOT

	Rank-1 Accuracy	VR @FAR = 10%	VR @FAR = 1%	VR @FAR = 0.1%
Cosine similarity	37.22%	76.07%	51.21%	29.67%
LDA	38.36%	86.65%	56.66%	30.50%
JB	41.64%	85.67%	61.75%	38.13%
HJB	**50.82%**	**94.74%**	**77.29%**	**55.10%**

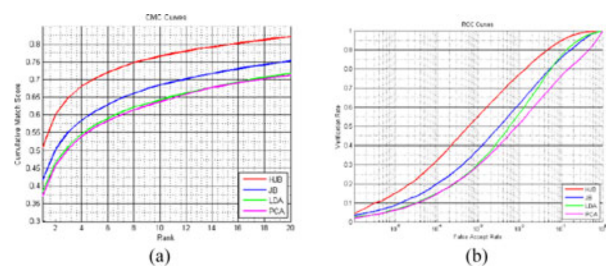

Fig. 3. (a) CMC and (b) ROC curves on the ID versus spot recognition task.

To perform the evaluation, we divide the dataset into ten subsets with nonoverlapping equal number of subjects in each. Then, the tenfold cross-validation is performed. In each fold, nine subsets are used for training, and the remaining one is used for test. We apply the features extracted by the model from [31] as the input for this experiment.

Table III lists the performance of different methods, whose corresponding CMC and ROC curves are shown in Fig. 3. Four methods, including cosine similarity, LDA, JB, and HJB, are compared. As expected, LDA, JB, and HJB, which learn discriminative metric, achieve higher face recognition performance than cosine similarity. The HJB, which models the modality differences, achieves the best and improves the JB by a large margin.

V. CONCLUSION

In this paper, we develop an asymmetric formulation from the JB model for heterogeneous face recognition. The modality difference is involved, so the HJB is more adaptive to cross-modality face matching. The metric is learned via optimizing the parameters in each modality separately. We evaluate the HJB on the benchmarks of CASIA-HFB and CASIA NIR-VIS 2.0 and obtain better results than the baseline JB and most of the other existing methods. The effectiveness of the HJB is also validated in the case of ID versus spot photo recognition.

REFERENCES

[1] S. Z. Li, S. R. Chu, S. Liao, and L. Zhang, "Illumination invariant face recognition using near-infrared images," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 29, no. 4, pp. 627–639, Apr. 2007.

[2] S. Liao, D. Yi, Z. Lei, R. Qin, and S. Z. Li, "Heterogeneous face recognition from local structures of normalized appearance," in *Advances in Biometrics*. New York, NY, USA: Springer, 2009, pp. 209–218.

[3] W. Zhang, X. Wang, and X. Tang, "Coupled information-theoretic encoding for face photo-sketch recognition," in *Proc. IEEE Conf. Comput. Vision Pattern Recognit.*, 2011, pp. 513–520.

[4] S. Liu, D. Yi, Z. Lei, and S. Z. Li, "Heterogeneous face image matching using multi-scale features," in *Proc. 5th IAPR Int. Conf. Biometrics*, 2012, pp. 79–84.

[5] Z. Lei, M. Pietikainen, and S. Z. Li, "Learning discriminant face descriptor," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 36, no. 2, pp. 289–302, Feb. 2014.

[6] D. Lin and X. Tang, "Inter-modality face recognition," in *Proc. 9th Eur. Conf. Comput. Vision*, 2006, pp. 13–26.

[7] Z. Lei, S. Liao, D. Yi, R. Qin, and S. Z. Li, "A discriminant analysis method for face recognition in heteroscedastic distributions," in *Advances in Biometrics*, New York, NY, USA: Springer, 2009, pp. 112–121.

[8] Z. Lei and S. Z. Li, "Coupled spectral regression for matching heterogeneous faces," in *Proc. IEEE Conf. Comput. Vision Pattern Recognit.*, 2009, pp. 1123–1128.

[9] Z. Lei, C. Zhou, D. Yi, A. K. Jain, and S. Z. Li, "An improved coupled spectral regression for heterogeneous face recognition," in *Proc. 5th IAPR Int. Conf. Biometrics*, 2012, pp. 7–12.

[10] B. Klare and A. K. Jain, "Sketch-to-photo matching: A feature-based approach," *Proc. SPIE*, vol. 7667, 2010, Art. no. 766702.

[11] B. F. Klare, Z. Li, and A. K. Jain, "Matching forensic sketches to mug shot photos," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 33, no. 3, pp. 639–646, Mar. 2011.

[12] Z. Li, D. Gong, L. Qiang, D. Tao, and L. Xuelong, "Mutual component analysis for heterogeneous face recognition," *ACM Trans. Intell. Syst. Technol.*, vol. 7, 2016, Art. no. 28.

[13] R. Wang, J. Yang, D. Yi, and S. Z. Li, "An analysis-by-synthesis method for heterogeneous face biometrics," in *Advances in Biometrics*. New York, NY, USA: Springer, 2009, pp. 319–326.

[14] F. Juefei-Xu, D. Pal, and M. Savvides, "NIR-VIS heterogeneous face recognition via cross-spectral joint dictionary learning and reconstruction," in *Proc. IEEE Conf. Comput. Vision Pattern Recognit. Workshops*, 2015, pp. 141–150.

[15] S. Yin, X. Dai, P. Ouyang, L. Liu, and S. Wei, "A multi-modal face recognition method using complete local derivative patterns and depth maps," *Sensors*, vol. 14, no. 10, pp. 19561–19581, 2014.

[16] A. Aissaoui, J. Martinet, and C. Djeraba, "DLBP: A novel descriptor for depth image based face recognition," in *Proc. IEEE Int. Conf. Image Process.*, 2014, pp. 298–302.

[17] Y. Sun, Y. Chen, X. Wang, and X. Tang, "Deep learning face representation by joint identification-verification," in *Proc. Adv. Neural Inf. Process. Syst. Conf.*, 2014, pp. 1988–1996.

[18] O. M. Parkhi, A. Vedaldi, and A. Zisserman, "Deep face recognition," in *Proc. Brit. Mach. Vision Conf.*, 2015, vol. 1, 2015, p. 6.

[19] D. Yi, Z. Lei, and S. Z. Li, "Shared representation learning for heterogenous face recognition," in *Proc. 11th IEEE Int. Conf. Workshops Autom. Face Gesture Recognit.*, 2015, vol. 1, pp. 1–7.

[20] Y. Jin, J. Cao, Y. Wang, and R. Zhi, "Ensemble based extreme learning machine for cross-modality face matching," *Multimedia Tools Appl.*, vol. 75, pp. 11831–11846, 2016.

[21] Y. Jin, J. Li, C. Lang, and Q. Ruan, "Multi-task clustering ELM for VIS-NIR cross-modal feature learning," *Multidimensional Syst. Signal Process.*, pp. 1–16, 2016, doi: 10.1007/s11045-016-0401-8.

[22] X. Liu, L. Song, X. Wu, and T. Tan, "Transferring deep representation for NIR-VIS heterogeneous face recognition," in *Proc. Int. Conf. Biometrics*, 2016, pp. 1–8.

[23] D. Chen, X. Cao, L. Wang, F. Wen, and J. Sun, "Bayesian face revisited: A joint formulation," in *Proc. 12th Eur. Conf. Comput. Vision*, 2012, pp. 566–579.

[24] P. N. Belhumeur, J. P. Hespanha, and D. J. Kriegman, "Eigenfaces vs. fisherfaces: Recognition using class specific linear projection," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 19, no. 7, pp. 711–720, Jul. 1997.

[25] P. Li, Y. Fu, U. Mohammed, J. H. Elder, and S. J. Prince, "Probabilistic models for inference about identity," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 34, no. 1, pp. 144–157, Jan. 2012.

[26] Z. Lei, S. Liao, A. K. Jain, and S. Z. Li, "Coupled discriminant analysis for heterogeneous face recognition," *IEEE Trans. Inf. Forensics Security*, vol. 7, no. 6, pp. 1707–1716, Dec. 2012.

[27] S. Li, D. Yi, Z. Lei, and S. Liao, "The CASIA NIR-VIS 2.0 face database," in *Proc. IEEE Conf. Comput. Vision Pattern Recognit. Workshops*, 2013, pp. 348–353.

[28] D. Yi, R. Liu, R. Chu, Z. Lei, and S. Z. Li, "Face matching between near infrared and visible light images," in *Proc. Int. Conf. Adv. Biometrics*, 2007, pp. 523–530.

[29] T. I. Dhamecha, P. Sharma, R. Singh, and M. Vatsa, "On effectiveness of histogram of oriented gradient features for visible to near infrared face matching," in *Proc. 22nd Int. Conf. Pattern Recognit.*, 2014, pp. 1788–1793.

[30] Y. Jin, J. Lu, and Q. Ruan, "Coupled discriminative feature learning for heterogeneous face recognition," *IEEE Trans. Inf. Forensics Security*, vol. 10, no. 3, pp. 640–652, Mar. 2015.

[31] D. Yi, Z. Lei, S. Liao, and S. Z. Li, "Learning face representation from scratch," *arXiv preprint arXiv:1411.7923*, 2014.