The CASIA NIR-VIS 2.0 Face Database

Stan Z. Li, Dong Yi, Zhen Lei and Shengcai Liao
Center for Biometrics and Security Research & National Laboratory of Pattern Recognition
Institute of Automation, Chinese Academy of Sciences (CASIA)
szli, dyi, zlei, scliao@cbsr.ia.ac.cn

Abstract

In recent years, heterogeneous face biometrics has attracted more attentions in the face recognition community. After published in 2009, the HFB database [7] has been applied by tens of research groups and widely used for Near infrared vs. Visible light (NIR-VIS) face recognition. Despite its success the HFB database has two disadvantages: a limited number of subjects, lacking specific evaluation protocols. To address these issues we collected the NIR-VIS 2.0 database. It contains 725 subjects, imaged by VIS and NIR cameras in four recording sessions. Because the 3D modality in the HFB database was less used in the literature, we don't consider it in the current version. In this paper, we describe the composition of the database, evaluation protocols and present the baseline performance of PCA on the database. Moreover, two interesting tricks, the facial symmetry and heterogeneous component analysis (HCA) are also introduced to improve the performance.

1. Introduction

Heterogeneous face biometrics is a hot research topic in recent years. Researchers mainly focus on the following three problems: "Viewed Sketch vs. VIS" [14], "Forensic Sketch vs. VIS" [4] and "NIR vs. VIS" [6]. The first two problems often arise from forensics and security applications, while the "NIR vs. VIS" research aims to supply an alternative way to solve the illumination problem of traditional VIS face recognition system. To support these researches, many excellent databases are collected, such as CUHK Face Sketch Database (CUFS) [16], CASIA HFB Database (HFB) [7] and so on. Among them, the HFB database is the most popular one for "NIR vs. VIS" research.

Although the target of the HFB database is to improve the performance of "NIR vs. VIS" face recognition, other significance brought by the HFB database are also emerged unexpectedly in other aspects, such as joint learning across domains, face anti-spoofing and so on. For example, inspired by the coupled spectral regression (CSR) in [5], [17] trains a scale robust pedestrian detector by joint learning from the samples with different scales. A multi-spectral face recognition system in NIR-VIS spectrums is adopted in an EU anti-spoofing project (Tabula Rasa) [1], in which the multi-spectral spoofing database is constructed based on the HFB database. While the NIR and VIS images in the HFB database have been widely used in many fields, the 3D images have only been used by a few papers, such as [18]. For this reason, we don't consider this modality in this version of the database.

The HFB database was collected in the beginning of 2007 and the procedure lasted for several days. The subjects in the database are mainly the students in CASIA. For these reasons, the face images of HFB database are homogeneous with respect to age, acquisition environment and other variations, which are too simple to simulate practical applications. When the HFB database was first published in [7], the number of subjects was 100, and then the number increased to 202 in 2010. Compared to traditional face databases, the scale is still too small for some learning algorithms, *e.g.,* for LDA it may cause SSS (small sample size) problem [2]. Another defect is that the protocols for performance evaluation in [7] are roughly described without detailed protocol lists, which are unspecific for researchers to follow.

To complement the disadvantages of the HFB database, we collect a larger database called CASIA NIR-VIS 2.0 database, in which the images are captured using the same device as the HFB database. Compared to HFB, NIR-VIS 2.0 has the following new features:

1. The number of subjects in the NIR-VIS 2.0 database is 725, which is 3 times more than the HFB database.
2. We define a group of specific protocols for performance evaluation. On the contrary, the protocols of the HFB database are unclear for performance comparison or for reproducing experimental results.
3. In the new database, the age distribution of the subjects

are broader, spanning from children to old people.

4. The face images are collected in four recording sessions, which are from 2007 to 2010.

Additionally, the NIR-VIS 2.0 database contains more variations in pose and facial expression, therefore we believe this new database is more close to the practical situations. We recommend researchers to use the NIR-VIS 2.0 database instead of the HFB database. We will build a website [1], like LFW [2], to publish the results.

The rest of the paper is structured as follows. Section 2 reviews the methods related to heterogeneous face recognition and the HFB database. Section 3 describes the composition and content of the database. Section 4 defines the specific protocols for performance evaluation and reporting. Section 5 presents the performances of baseline methods: PCA [15] and its combination with two tricks. Section 6 summarizes the paper.

2. Literature Survey

In practical face recognition systems, face images may be captured in more than one modality. For example, the NIR based face recognition method [8] has been developed to overcome the illumination variation problem; sketch images drawn by artists based on the recollection of an eyewitness have been used in the retrieval of a sketch from the police mugshot databases. Therefore, heterogeneous face recognition is a current topic of interest. Different from traditional face recognition, the difficulty in heterogeneous face recognition mainly comes from the appearance differences between face images of different modalities. There are three kinds of methods to reduce such difference: face synthesis, invariant feature extraction and common subspace learning. Next, we will review the related work from the three aspects.

Early methods tended to synthesize one type of face image from another type, *e.g.,* synthesize photo using face sketch, and applied traditional face recognition methods on the synthesized face images. Representative works in this category include [14], [11] and [16], where the most representative one is eigen-transform. The eigen-transform was proposed by Tang and Wang in [13] for matching sketch images with photos. For a photo, it first computes the reconstruction coefficients using the photo training set. Subsequently, the same combination coefficients are used to synthesize pseudo sketch image with the corresponding sketch training images. Finally, the pseudo sketches are used for face recognition. Their method has been proved effective in reducing the difference between photo and sketch.

[1] http://www.cbsr.ia.ac.cn/english/NIR-VIS-2.0-Database.html

[2] http://vis-www.cs.umass.edu/lfw/

In the second category, researchers try to extract invariant features from heterogeneous face images. Proper texture descriptors are designed and applied to the heterogeneous images to reduce the difference between them. Liao *et al.* [9] first utilized the difference of Gaussian (DoG) filter to process the NIR and VIS images to reduce the appearance difference and then extracted multi-block local binary pattern (MBLBP) to represent faces. Klare and Jain [3] used HoG and LBP descriptors and learnt an ensemble of discriminant projections. In [4], the authors proposed to extract SIFT and multi-scale local binary patterns (MLBP) features from forensic sketches and mug shot photos, respectively. Zhang *et al.* [21] proposed a learning based coupled information theoretic encoding descriptor to capture a discriminant local structure for photo-sketch images and applied PCA+LDA classifier to compute the dissimilarity of samples. All the above methods try to reduce the gap between heterogeneous face images at the feature level and then apply traditional face classification methods to realize the recognition task.

Subspace learning methods are very popular for traditional face recognition. Because of its simplicity and success, subspace learning was naturally extended to heterogeneous face recognition by many researchers. The basic idea is to find a common discriminative subspace in which the representations of heterogeneous images from the same person are as close as possible while the representations of heterogeneous images from different persons are as far as possible. Lin and Tang [10] proposed a common discriminant feature extraction (CDFE) method to transform query faces captured using near infrared or sketch images and target faces of visible spectrum into a common discriminant feature subspace, where the ratio of between scatter matrix to within scatter matrix is maximized. Although CDFE achieves high recognition rate on training set, its generalization performance is poor. Yi *et al.* [20] utilized canonical correlation analysis (CCA) to exploit the essential correlations in PCA and LDA subspaces of NIR and VIS images and Yang *et al.* [18] proposed regularized kernel CCA to learn the relationship between VIS and 3D face data spaces. Lei and Li [5, 6] proposed the coupled spectral regression (CSR) method to deal with heterogeneous face recognition problem and achieved a better generalization performance than previous methods.

To our knowledge, there are five existing heterogeneous face databases: CUHK Face Sketch Database (CUFS) [16], CUHK Face Sketch FERET Database (CUFSF) [21], MSU Forensic Sketch Database [4], CASIA HFB Database (HFB) [7], and Long Distance Heterogeneous Face Database (LDHF-DB) [12]. Except for the MSU Forensic Sketch Database, other databases are all publicly available. Their information are shown in Table 1.

Figure 1. VIS and NIR face images, with variations in resolution, lighting conditions, pose and age, of one subject in the NIR-VIS 2.0 database.

3. Database Description

The images in the NIR-VIS 2.0 database were collected in four recording sessions: 2007 spring, 2009 summer, 2009 fall and 2010 summer, in which the first session is identical to the HFB database. In summary, the NIR-VIS 2.0 database consists of 725 subjects in total. There are 1-22 VIS and 5-50 NIR face images per subject. Figure 1 shows some face images of a subject in the database. Table 1 shows its detail and list the information of available heterogeneous face databases for reference.

The NIR-VIS 2.0 database includes the following contents:

1. The raw images, including the VIS images in JPEG format and the NIR images in BMP format. Their resolutions are both 640×480.

2. The eye coordinates of the VIS, NIR images. They are automatically labeled by an eye detector [19], and several error coordinates are corrected manually.

3. Cropped versions of the raw VIS, NIR images. The resolution is 128×128, and the process is done based on the eye coordinates.

4. Protocols for performance evaluation. The protocols include two views: one view for algorithm development, another for performance reporting.

Some samples are shown in Figures 1 and 2. The cropped VIS and NIR faces will be used for baseline experiments (see Section 5).

4. Evaluation Protocols

In order to report unbiased performance, we build two views from the database, in which View1 is for algorithm development and View2 is for performance reporting. The

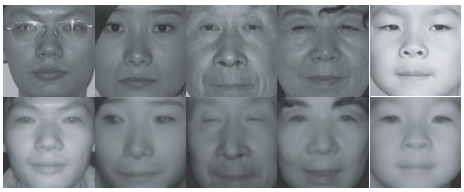

Figure 2. Cropped VIS (top row) and NIR (bottom row) face images from the NIR-VIS 2.0 database. Each column represents one person.

View2 includes ten sub-experiments. Parameters of algorithms are only allowed to be tuned on View1. While reporting performance on View2, the parameters must be fixed.

The image lists of the two views are all generated in random way and stored in several text files. The details are shown as follows:

1. View1: vis-train-dev.txt and nir-train-dev.txt are the image lists for training; vis-gallery-dev.txt and nir-probe-dev.txt are for testing.

2. View2: vis-train-[1-10].txt and nir-train-[1-10].txt are for training; vis-gallery-[1-10].txt and nir-probe[1-10].txt are for testing, where [1-10] denotes the id of sub-experiment.

The subjects in the training and the corresponding testing set are non-overlapping, and the percentage of subjects in the training and testing set are both nearly 50%. To simulate the practical situations, the VIS images are used as gallery and the NIR images are used as probe. For each subject in the gallery set, only one VIS image is selected.

Based on the protocols, we can tune parameters of algorithms on View1 and calculate the similarity (or distance) matrix of each sub-experiment on View2. The ROC curve and rank1 recognition rate are used to evaluate the final performance. The ROC curve is generated by all similarity scores and masks matrices of ten sub-experiments. For the rank1 recognition rate, the mean accuracy and standard deviation of ten sub-experiments should be reported. The mean accuracy is given by

$$\mu = \frac{\sum_{i=1}^{10} r_i}{10}, \tag{1}$$

where r_i is the rank1 recognition rate of the ith sub-experiment on View 2. The standard deviation is given by

$$\sigma = \sqrt{\frac{\sum_{i=1}^{10}(r_i - \mu)^2}{9}}. \tag{2}$$

Note that the cropped 128×128 face images in the database are just for convenience. Researchers are encouraged to use other sizes or alignment methods to improve the recognition rate.

Table 1. Summary of publicly available databases and the proposed database for heterogeneous face recognition. The last column denotes the main variation they addressed, *i.e.,* Pose(P), Expression(E), Eyeglasses(G), Distance(D).

Name of Database	Modalities	No. of Subjects	No. of Images	Variations
CUFS	Sketch, VIS	606	1212	-
CUFSF	Sketch, VIS	1194	2388	-
LDHF-DB	NIR, VIS	100	1600	D
CASIA-HFB	NIR, VIS	202	5097	E, G, D
CASIA NIR-VIS 2.0	NIR, VIS	725	17580	P, E, G, D

5. Baseline Performance

To illustrate the usage of the database, PCA is first used as baseline method. And then we introduce two tricks to improve its performance: (1) facial symmetry is used to augment the dataset and improve the computation efficiency; (2) hetero-component analysis is used to remove the difference between NIR and VIS face images. The two tricks may contribute to other advanced methods in the future.

5.1. PCA

Before experiments, all cropped 128×128 face images are normalized to zero mean and unit length. For PCA, the NIR and VIS face images are mixed together to train a subspace, the dimension of which is retained by preserving the 98% of the total energy. In the testing phase, the Cosine metric (see Equation 3) is adopted to evaluate the similarity of samples and the maximum correlation criterion is used for classification.

$$s(\mathbf{x},\mathbf{y}) = \frac{(\mathbf{x}-\mathbf{m})^T(\mathbf{x}-\mathbf{m})}{\sqrt{(\mathbf{x}-\mathbf{m})^T(\mathbf{x}-\mathbf{m})(\mathbf{y}-\mathbf{m})^T(\mathbf{y}-\mathbf{m})}}, \quad (3)$$

where $\mathbf{x}$ and $\mathbf{y}$ are two face samples, and $\mathbf{m}$ is the mean of training set.

The result of PCA and the following methods are all-shown in Figure 6 and Table 2. From the results, we can see that the performance of PCA on the HFB database [7] (refer to Table 4, Exp. 1) and the current database are both poor. This illustrates that PCA is originally proposed to represent single modal distribution (or Gaussian), which is not appropriate for multi-modal problem.

5.2. Facial Symmetry

According to the protocol, the number of samples in the training set of View1 is about 8600, which is smaller than the dimension of face sample (128×128 = 16384). Therefore, the training set can not fully span the space of face image, thus degrading the performance and generalization ability of PCA.

Facial symmetry is a useful cue in face analysis, and has been widely explored for face synthesis, face detection, pose robust face recognition and so on. Here, we use the facial symmetry cue to reduce the dimension of samples and meanwhile augment the number of samples. For all face images, we divide them into two halves and mirror the right half to left, then the dimension is reduced to $128 \times 64 = 8192$ and the number of samples is multiplied 2. The process is shown in Figure 3.

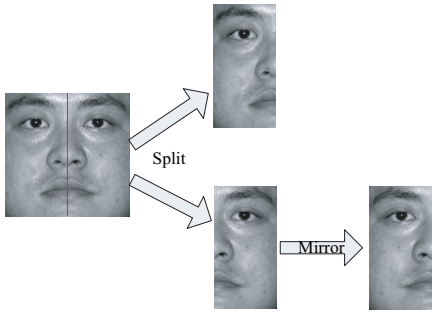

Figure 3. Splitting a face sample into two samples by the facial symmetry, which not only augments the number of samples but also reduces the dimensionality.

From the results in Figure 6 and Table 2, we can see the performances of PCA is improved by this trick, *e.g.,* the mean accuracy is improved from 7.16% to 9.26%.

5.3. Hetero-Component Analysis

Different from traditional face recognition, the distribution of NIR and VIS images are multi-modal in the image (or feature) space. The multi-modality is a key property of heterogenous face recognition. We can assert that how to deal with or how to build connection between the two modes is the key to solve the problem. A toy example is shown in Figure 4 to illustrate the distribution of NIR and VIS images, where the blue points denote the VIS images; the red points denote the NIR images; and the black line is the direction from the center of VIS images to that of NIR images, which is called the hetero-component. Because the difference between NIR and VIS mode usually dominant the difference caused by other factors, such as identity, illumination, expression and so on, the hetero-component is usually related with the first component of PCA. By removing the component, we can merge the NIR and VIS modes together (Figure 4). Although this operation can decrease the difference between modalities, meanwhile, its disadvantages should be studied in the future.

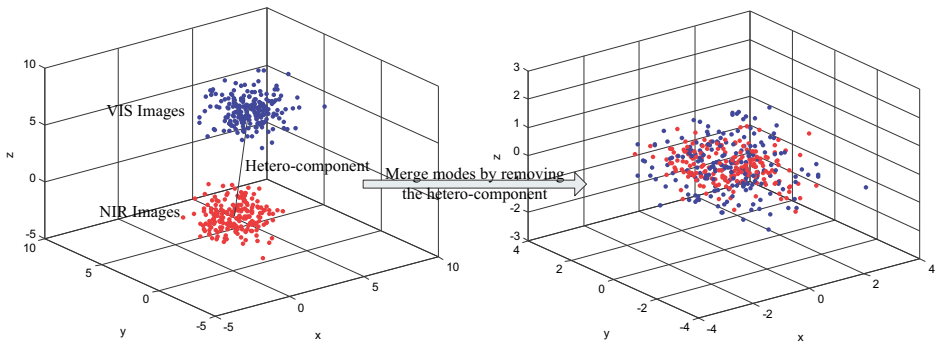

Figure 4. A toy example to illustrate the multi-modal distributions of VIS and NIR images, and the merged distribution after removing the Hetero-component.

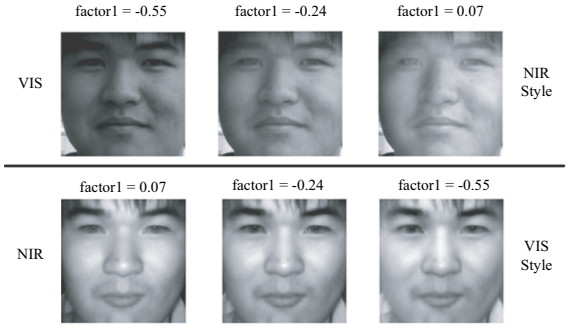

Figure 5. Transforming between VIS and NIR face images by manipulating the factor of the first PCA component.

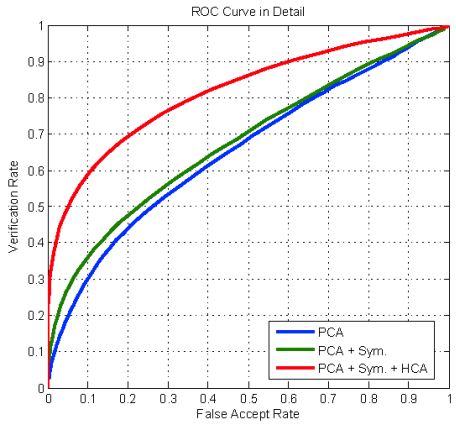

Figure 6. The ROC curves of PCA and its variants on View2 of the NIR-VIS 2.0 database.

In the experiment, we evaluate the correlation between the first component of PCA and the hetero-component by

$$\rho = \frac{\mathbf{p}_1^T \mathbf{h}}{\sqrt{\mathbf{h}^T \mathbf{h}}} \tag{4}$$

where $\mathbf{p}_1$ is the first component of PCA; $\mathbf{h}$ is the hetero-component calculated by $\mathbf{h} = \mathbf{m}_{NIR} - \mathbf{m}_{VIS}$. The correlations on the 10 sub-experiments of View2 are all around 0.97, which shows that PCA can capture the heterogeneity of face images in the first component.

To illustrate the property of the first component of PCA more clearly, we conduct a face synthesis experiment on View1. First, we build a PCA subspace using the training set, then project a VIS image, in the testing set, into the subspace. By tuning the factor of the first component, we find the VIS image can transform into an NIR-style image, which is shown in Figure 5. When we do the same process on an NIR image, it will transform into a VIS-style image too. Two examples in Figure 5 verify that the first component exactly capture the heterogeneity caused by the spectrum of illumination.

Inspired by some PCA related works, the first several components of PCA, besides the first one, also can capture variations in illumination, pose and occlusion, therefore we could remove the first several components of PCA to improve the performance by Equation 5.

$$\mathbf{x}' = \mathbf{x} - \mathbf{w}_k \mathbf{w}_k^T \mathbf{x}, \tag{5}$$

where $\mathbf{x}$ is the original sample; $\mathbf{w}_k$ is the the first k components of PCA; $\mathbf{w}_k \mathbf{w}_k^T \mathbf{x}$ can be seen as the heterogeneous part of the sample. In experiments, the optimal number of k, tuned on View1, is 25. We call this trick as heterogeneous component analysis (HCA) in the current context.

The results of HCA, combing with facial symmetry, are also shown in Figure 6 and Table 2. After using the HCA trick, the ROC curve and rank1 recognition rate are all improved significantly. And we expect this trick can also help to improve other state-of-the-art feature based methods. Similar to the DoG filter, enhancing the mid-frequency information in image space, HCA is a analogy of mid-pass filter in PCA subspace.

Table 2. The rank1 recognition rate of PCA and its variants on View2 of the NIR-VIS 2.0 database.

Methods	Mean accuracy	Std. deviation
PCA	7.16%	0.52%
PCA + Sym.	9.26%	0.66%
PCA + Sym. + HCA	23.70%	1.89%

6. Summary

In this paper, we introduced a larger scale, more realistic NIR-VIS database than existing NIR-VIS face databases. To make up the disadvantages of the HFB database, the CASIA NIR-VIS 2.0 database contains 3 times more subjects and the face images are captured with more variations. The biggest highlight of the proposed database is its detailed evaluation protocols, using which researchers can easily generate reproducible and comparable results. In experiments, the performance of PCA is reported as baseline, and two tricks are further used to improve the performance. Especially, the HCA is an interesting clue to analyze the relationship between NIR and VIS face images. In the future, we will collect more face images in more spectrums, besides VIS and NIR, which may open a way to study the property of face image along the spectral dimension.

Acknowledgements

This work was supported by the NSFC Project #61070146, #61105023, #61103156, #61105037, #61203267, National IoT R&D Project #2150510, National Science and Technology Support Program Project #2013BAK02B01, EU FP7 Project #257289 (TABULA RASA), and AuthenMetric R&D Funds.

References

[1] Trusted biometrics under spoofing attacks. http://www.tabularasa-euproject.org/. 1

[2] P. N. Belhumeur, J. P. Hespanha, and D. J. Kriegman. "Eigenfaces vs. Fisherfaces: Recognition using class specific linear projection". In *Proceedings of the European Conference on Computer Vision*, pages 45–58, 1996. 1

[3] B. Klare and A. Jain. "Heterogeneous face recognition: Matching NIR to visible light images". In *International Conference on Pattern Recognition*, pages 1513–1516, 2010. 2

[4] B. Klare, Z. Li, and A. Jain. "Matching forensic sketches to mug shot photos". *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, 33(3):639–646, 2011. 1, 2

[5] Z. Lei and S. Li. "Coupled spectral regression for matching heterogeneous faces". In *IEEE Conference on Computer Vision and Pattern Recognition*, pages 1123–1128, 2009. 1, 2

[6] Z. Lei, S. Liao, A. Jain, and S. Li. "Coupled discriminant analysis for heterogeneous face recognition". *IEEE Transactions on Information Forensics and Security*, 7(6):1707–1716, 2012. 1, 2

[7] S. Li, Z. Lei, and M. Ao. "The HFB face database for heterogeneous face biometrics research". In *IEEE Conference on Computer Vision and Pattern Recognition Workshops*, pages 1–8, 2009. 1, 2, 4

[8] S. Z. Li, R. Chu, S. Liao, and L. Zhang. "Illumination invariant face recognition using near-infrared images". *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 26, April 2007. 2

[9] S. Liao, D. Yi, Z. Lei, R. Qin, and S. Li. "Heterogeneous face recognition from local structures of normalized appearance". In M. Tistarelli and M. Nixon, editors, *Advances in Biometrics*, volume 5558 of *Lecture Notes in Computer Science*, pages 209–218. 2009. 2

[10] D. Lin and X. Tang. "Inter-modality face recognition ". In *Proceedings of the European Conference on Computer Vision*, volume 3954, pages 13–26, 2006. 2

[11] Q. Liu, X. Tang, H. Jin, H. Lu, and S. Ma. "A nonlinear approach for face sketch synthesis and recognition". In *Proceedings of IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, volume 1, pages 1005–1010, 2005. 2

[12] H. Maeng, S. Liao, D. Kang, S.-W. Lee, and A. K. Jain. "Nighttime face recognition at long distance: Cross-distance and cross-spectral matching". In *Proceedings of the First Asian Conference on Computer Vision*, pages 5–9, Daejeon, Korea, 2012. 2

[13] X. Tang and X. Wang. "Face sketch synthesis and recognition". In *IEEE International Conference on Computer Vision*, volume 1, pages 687–694, 2003. 2

[14] X. Tang and X. Wang. "Face sketch recognition". *IEEE Transactions on Circuits and Systems for Video Technology*, 14(1):50–57, Jan. 2004. 1, 2

[15] M. A. Turk and A. P. Pentland. "Face recognition using eigenfaces". In *Proceedings of IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, pages 586–591, Hawaii, June 1991. 2

[16] X. Wang and X. Tang. "Face photo-sketch synthesis and recognition". *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, 31(11):1955–1967, 2009. 1, 2

[17] J. Yan, X. Zhang, Z. Lei, D. Yi, S. Liao, and S. Z. Li. "Robust multi-resolution pedestrian detection in traffic scenes". In *IEEE Conference on Computer Vision and Pattern Recognition*, 2013. 1

[18] W. Yang, D. Yi, Z. Lei, J. Sang, and S. Li. "2D-3D face matching using cca". In *IEEE International Conference on Automatic Face Gesture Recognition*, pages 1–6, 2008. 1, 2

[19] D. Yi, Z. Lei, and S. Z. Li. "A robust eye localization method for low quality face images". In *International Joint Conference on Biometrics (IJCB)*, pages 15–21, Washington, DC, USA, Oct. 11-13 2011. 3

[20] D. Yi, R. Liu, R. Chu, Z. Lei, and S. Z. Li. "Face matching between near infrared and visible light images". In *Proceedings of IAPR International Conference on Biometric*, Seoul, Korea, August 2007. 2

[21] W. Zhang, X. Wang, and X. Tang. "Coupled information-theoretic encoding for face photo-sketch recognition". In *IEEE Conference on Computer Vision and Pattern Recognition*, pages 513–520, 2011. 2