The HFB Face Database for Heterogeneous Face Biometrics Research

Stan Z. Li, Zhen Lei, Meng Ao
Center for Biometrics and Security Research
Institute of Automation, Chinese Academy of Sciences
95 Zhongguancun Donglu, Beijing 100190, China
{szli, zlei, mao}@cbsr.ia.ac.cn

Abstract

A face database, composed of visual (VIS), near infrared (NIR) and three-dimensional (3D) face images, is collected. Called the **HFB Face Database***, it is released now to promote research and development of Heterogeneous Face Biometrics (HFB). This release of Version 1 contains a total of 992 images from 100 subjects; there are 4 VIS, 4 NIR, and 1 or 2 3D face images per subject. In this paper, we describe the apparatuses, environments and procedure of the data collection and present baseline performances of the standard PCA and LDA methods on the database.*

1. Introduction

Face images can be captured in different spectral bands, *e.g.* , in Visual (VIS), near infrared (NIR), or thermal infrared (TIR), or as measurements of 3D facial shape. These different image types, due to different image formation characteristics or principles, are said to be *heterogeneous*. Although heterogeneous face images of a given person differ by pixel values, the identify of the face should be classified as the same. Biometric methods by processing and matching between heterogeneous face images are collectively referred to as **heterogeneous face biometrics (HFB)** [11].

While face biometrics have traditionally been based on VIS face images, recent years we have seen developments in non-visual image based face biometrics, including 3D [4], TIR [8], and NIR [12] based methods. These methods are advantageous over VIS image based ones in that they are less affected by illumination changes. However, on the other hand, many applications, such as E-Passport, require visual (VIS) image based enrollment.

Interest in HFB arises from applications involving matching between heterogeneous face images. For example, one would like to take NIR face images, which overcome uncontrolled illumination changes, as the query, and match them against VIS face images in the target set as required by applications. Study on relationship and mapping between heterogeneous face images is also an interesting problem from pattern recognition and machine learning viewpoints.

The latest multiple biometric grand challenge (MBGC 2008) [17] has set up a series of experiments to evaluate recent advances of biometric technologies. In the MBGC portal challenge, NIR face videos are used as one of the probe sets whereas the target set consists of full faces in VIS images.

HFB by matching face images from different image sources has become a new direction for face recognition research. The earliest HFB could be face identification based on face sketches in forensic investigations. Tang and his colleagues developed a PCA based method for face sketch recognition against face photos [22]. In [15], the authors proposed a framework for inter-modality face matching called common discriminant feature extraction (CDFE). CDFE can be considered as a modified linear discriminant analysis (LDA) method for heterogeneous face matching, in which scatter matrices are defined on two different types of images, and local consistency of manifold is imposed to regularize the dimension reduction. It was applied to sketch-photo image matching and NIR-VIS face image matching in [15].

Stan Li's group published a series of papers on HFB research in recent years. In [27], a method was proposed for NIR-VIS face image matching problem based on canonical correlation analysis (CCA) [7]. Given that NIR-VIS image pairs of subjects are available for training, a CCA based correlation mechanism was learned from corresponding NIR-VIS face images. The CCA learning was performed in some feature subspaces such as PCA or LDA, rather than in the image space, and better classification performance was gained. In [26], LoG features were extracted and a non-learning based method was proposed for matching NIR and VIS face images in the MBGC portal challenge [17]. In [13], DoG [21] preprocessing was adopted

to alleviate the differences of NIR and VIS images, and R-LDA [16] combined with MB-LBP features [14] was then utilized to derive the most discriminant subspace. In [10], a coupled spectral regression framework was proposed to match between NIR and VIS face images. In [9], a CCA based regression method, that worked in NIR and 3D tensor spaces, was proposed to recover 3D face shape from a single NIR face image. In [25], a regularized kernel CCA and patch strategy was proposed to learn relationship between VIS and 3D face images. In [24], an analysis-by-synthesis method is developed to transform an NIR face query image to its VIS counterparts, and then compared to the target template VIS face images.

In this paper, we describe a face database, called the "HFB Face Database", for HFB research. Visual (VIS), near infrared (NIR) and three-dimensional (3D) face images are collected. The HFB Face Database is produced and released to promote and advance HFB research and development. This release of Version 1 contains a total of 992 images from 100 subjects. For every subjects, there are 4 VIS, 4 NIR, and 1 or 2 3D face images. We present the the baseline performances of the standard PCA and LDA methods for the database.

Several face databases have been existing, including VIS databases such as AT&T [1], CMU [20], FERET [19], FRGC [18] etc., CBSR NIR database [5], and 3D databases like FRAV3D [6], BJUT-3D [3] and so on. The HFB database described in this paper will add a new dimension in face recognition research.

In the rest of the paper, Section 2 introduces the devices used to collect different modalities of faces. Section 3 and 4 depict the collection procedure and content of the database. The baseline performances with different experiment configurations are presented in Section 5.

2. Data Acquisition Setup

This section describes the hardware devices and environments for the collection of the VIS, NIR and 3D face images.

2.1. Visual Face Images

The visual color face images are captured using Canon A640 Camera. The image resolution is 640×480. The camera is fixed on a tripod and the subjects are asked to be seated on a chair in front of the white background, as shown in Fig. 1. The height of digital camera is adjusted manually so that the face is in the central part of the image. Four frontal images with neutral and smile expression (or with and without glasses) as well as two different distances are captured.

Figure 1. The VIS camera and the capture environment.

2.2. NIR face images

The NIR images are captured by a home-brew device [12]. In order to obtain good quality NIR image, two strategies are designed to control the light direction: (1) mount active lights on the camera to provide frontal lighting, (2) minimize environmental lighting. We set two requirements on the active lighting: (1) the lights should be strong enough to produce a clear frontal-lighted face image without causing disturbance to human eyes, and (2) the minimization of the environmental lighting should have minimum reduction of the intended active lighting.

We use NIR light-emitting diodes (LEDs) of 850nm as the active lighting source, which are strong enough for indoor use and are power-effective. Such NIR lights are almost invisible to human eyes, yet most CCD and CMOS sensors have sufficient response at this spectrum point.

When mounted on the camera, the LEDs are approximately co-axial to the camera direction, and thus provide the best possible straight frontal lighting, better than mounted anywhere else; moreover, when the LEDs and camera are together, control of the lights can be more easily using a circuit in the box. The geometric layout of the LEDs on the camera panel may be carefully designed such that the illumination on the face is as homogeneous as possible.

Figure 2. The NIR face camera and the capture environment.

The strength of the total LED lighting should be such that it results in the NIR face images with good S/N ratio when the camera-face distance is between 80-120 cm, a conve-

nient range for the user. A guideline is that it should be as strong as possible, at least stronger than expected environmental illumination, yet does not cause sensor saturation. A concern is the safety to human eyes. When the sensor working in the normal mode is not saturated, the safety is guaranteed.

Our solution for requirement (2) above is to use a long pass optical filter to cut off visible light while allowing NIR light to pass. We choose a filter such that ray passing rates are 0%, 50%, 88%, and 99% at the wavelength points of 720, 800, 850, and 880nm, respectively. The filter cuts off visible environmental lights ($<$ 700nm) while allowing most of the 850nm NIR light to pass.

Fig. 2 illustrates the hardware device and its relationship with the face. The device consists of 18 NIR LEDs, an NIR camera, and the box. The NIR LEDs and camera are for NIR face image acquisition. The hardware and the face are relatively positioned in a way that the lighting is frontal and NIR rays provide nearly homogenous illumination on face. The imaging hardware works at a rate of 30 frames per second with the USB 2.0 protocol for 640×480 images.

The NIR imaging device used in this paper is a standard version for indoor use and under fluorescent lighting. For use in conditions with significant NIR components, it would require an enhanced version. The enhanced version developed in [28] provides clear face images even under sunlight.

The subjects are seated on a chair facing the NIR capture device. Subjects are asked to change their expression and move near and far from the camera. The NIR device captures a series of images of the process and we finally select four images with different expressions and poses.

2.3. 3D Face Images

The 3D faces are acquired using a Minolta vivid 910 laser scanner, as shown in Fig. 3. The laser scanner provides the depth (z-direction) for every point in the visible (x,y) plane of the faces. The background color is set to be black in order to absorb the laser better to reduce the noise of the captured data.

The process of 3D face acquisition takes about 3 seconds per scan. The subjects are asked to tidy their hair so as not to block the eyebrows and eyes, and keep still during the acquisition process. For most subjects, two 3D faces with different distances are acquired.

3. Database Description

Thus, the HFB face database consists of 100 subjects in total, including 57 males and 43 females. There are 4 VIS and 4 NIR face images per subject. For the 3D faces, there are 2 face images per subject for 92 subjects and 1 per subject for the other 8 subjects. Fig. 4 shows some face image examples.

Figure 3. 3D face capture device and environment. Subjects are seated in front of black background and scanned by Minolta vivid 910 laser scanner.

This release of HFB database Ver.1 includes the following:

1. The raw images, including (1) the VIS and (2) the NIR images of size 640×480 in the JPEG format, and (3) the 3D faces with wrl format;

2. The processed 3D faces: Face regions not belonging to the face are discarded and then the 3D data is processed by removing high noise and filling holes using an interpolation algorithm. Finally, the 3D points are sampled to form the depth images. The sizes vary from image to image.

3. The eye coordinates of the VIS, NIR and 3D depth images, manually labeled.

4. Cropped versions of the raw VIS, NIR and depth images. The crop was made in two sizes, 32×32 and 128×128, and is done based on the eye coordinates.

Some samples are shown in Figs. 4 and 5. The cropped VIS, NIR and depth faces are used for baseline experiments (see the next section). The download website is http://www.cbsr.ia.ac.cn/english/Databases.asp.

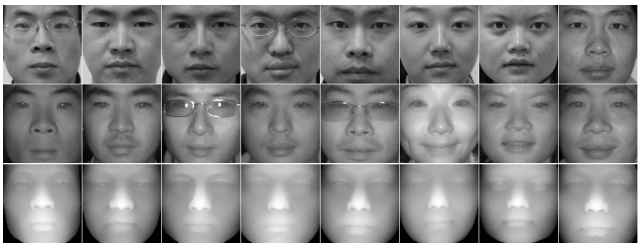


Figure 5. Cropped face examples of VIS (1st row), NIR (2nd row) and 3D depth faces (3rd row).

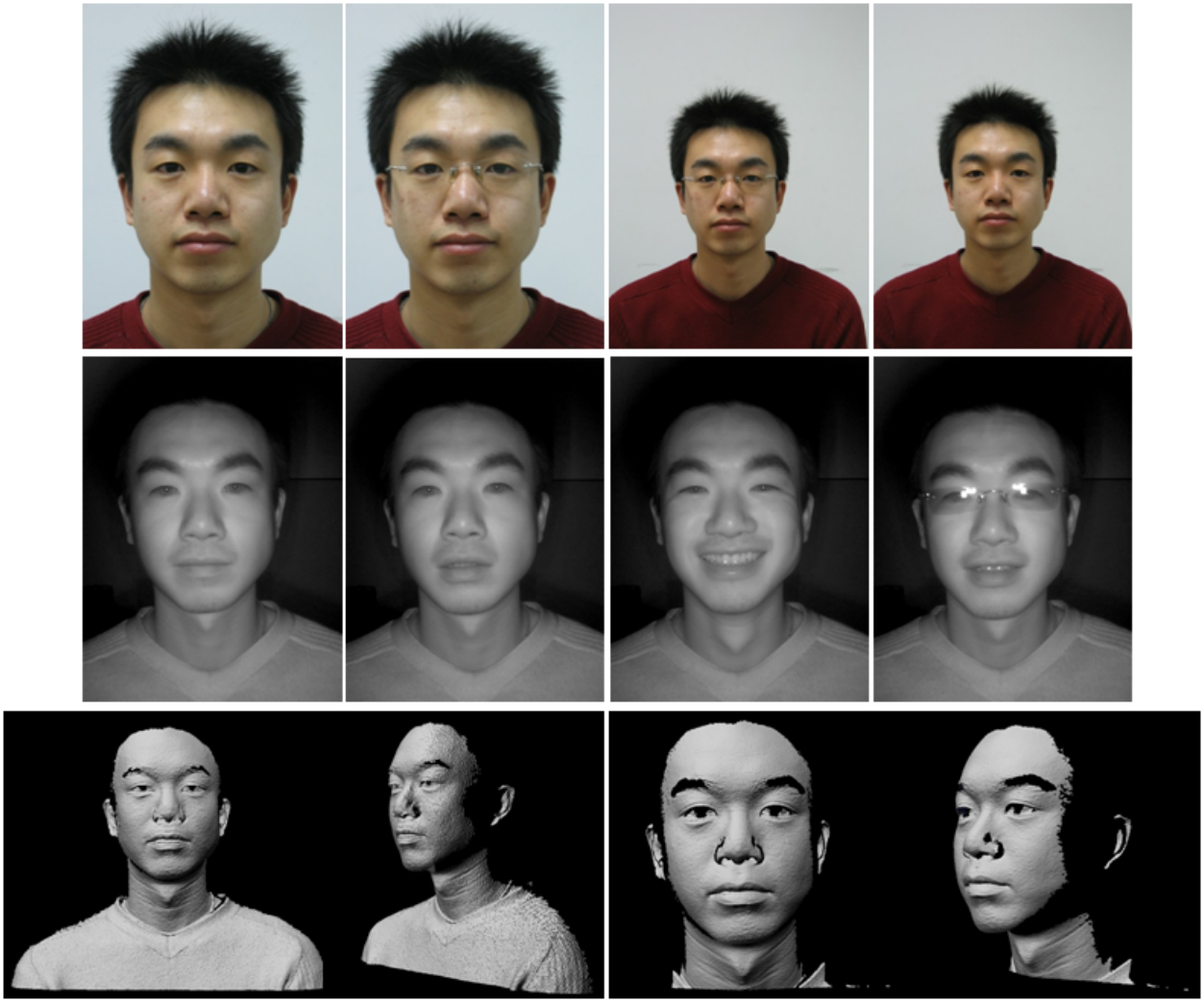


Figure 4. VIS, NIR and 3D raw data from one subject.

4. Baseline Results

The standard PCA [23] and LDA [2] are used as the baseline methods. Six experiments, explained in table 1, are designed to obtain the baseline performance on the cropped VIS, NIR and 3D depth images.

Table 1. Configurations of six experiments.

	Exp.1	Exp.2	Exp.3	Exp.4	Exp.5	Exp.6
Gallery	VIS	NIR	VIS	3D	NIR	3D
Probe	NIR	VIS	3D	VIS	3D	NIR

Two protocols are designed to test the baseline performances for every of the six experiments. In protocol I, 2 VIS and 2 NIR images are selected per subject, and this forms the training set; and the remaining 2 VIS and 2 NIR images are used for the test set. For 3D faces, for the subjects containing two images, 1 is selected for training and the other is used for testing. For the subjects containing only 1 image, the image is selected for testing. There is no overlapping in images between training and testing set. In protocol II, the images of the first 70 subjects are selected to form the training set, and the images of the last 30 subjects are used as the testing set. There is no overlapping in subjects between training and testing sets.

For the PCA method, the dimension of subspace is retained by preserving the 98% of the total energy. Table 2 illustrates the reduced PCA dimensions for the experiments. For the LDA method, more specifically, the Fisher-LDA method [2], PCA is first performed to make the with-in class scatter matrix non-singular, giving a subspace of dimensionality of sample_num minus class_num; LDA is then applied on the reduced subspace to find the final $c-1$ LDA dimensions, where c is the class number. Therefore, the final dimensionality of the LDA feature is 99 and 69 for the experiments in Protocol I and Protocol II, respectively. In

the testing phase, the cosine distance is adopted to evaluate the similarity of different samples and the maximum correlation criterion is used for the classification.

Table 2. Reduced dimension of PCA for different experiments.

	Protocol I		Protocol II	
	32x32	128x128	32x32	128x128
Exp.1&Exp.2	75	149	78	175
Exp.3&Exp.4	51	92	46	93
Exp.5&Exp.6	43	84	49	103

Tables 3 and 4 show the rank-1 recognition rates for six experiments following protocol I and II with two size images respectively and Figs. 6 and 7 illustrate the corresponding cumulative match curves.

We have the following conclusions for the baseline performances:

1. For the HFB problems in the baseline experiments, the unsupervised method of the PCA performed poorly. Comparatively, the supervised method of the Fisher-LDA performed much better.

2. For the LDA, the performance of protocol I was much better than that of protocol II, due to the fact that the subjects are all seen in both training and testing sets. However, Protocol II is often more relevant to practical situations in applications. The ability to generalize is a great challenge in HFB.

3. Among the six experiments, the results of Exp. 1 and Exp. 2 were better than others. That was because the imaging formation principles are relatively similar for VIS and NIR images, but are completely different from that of 3D depth imaging.

4. In summary, the baseline performance of the HFB face database was relatively poor. Development of effective methods are needed to create high performance HFB algorithms .

5. Conclusions

In this paper, we introduced the HFB face database of VIS-NIR-3D face images for HFB research. The imaging devices, environments and acquisition procedures are described. Six experiments designed for the baseline performances on the database are presented. In the future, we will collect more face images in several sessions, to provide time elapse variations and to enlarge the HFB database.

Acknowledgements

This work was supported by the following funding resources: National Natural Science Foundation Project #60518002, National Science and Technology Support Program Project #2006BAK08B06, National Hi-Tech (863) Program Projects #2006AA01Z192, #2006AA01Z193, and #2008AA01Z124, Chinese Academy of Sciences 100 people project, and AuthenMetric R&D Funds.

References

[1] AT&T Database. http://www.cl.cam.ac.uk/research/dtg/attarchive/facedatabase.html. 2

[2] P. Belhumeur, J. Hespanha, and D. Kriegman. Eigenfaces vs. fisherfaces: recognition using class specific linear projection. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 19(7):711–720, 1997. 4

[3] BJUT-3D Database. http://www.bjpu.edu.cn/sci/multimedia/mul-lab/3dface/facedatabase.htm. 2

[4] K. W. Bowyer, Chang, and P. J. Flynn. A survey of 3D and multi-modal 3d+2d face recognition. In *Proceedings of International Conference on Pattern Recognition*, pages 358–361, August 2004. 1

[5] CBSR NIR Face Dataset. http://www.cse.ohio-state.edu/otcbvs-bench/. 2

[6] FRAV3D Database. http://www.frav.es/databases/FRAV3d/. 2

[7] H. Hotelling. Relations between two sets of variates. *Biometrika*, 28:321–377, 1936. 1

[8] S. G. Kong, J. Heo, B. Abidi, J. Paik, and M. Abidi. Recent advances in visual and infrared face recognition - A review. *Computer Vision and Image Understanding*, 97(1):103–135, January 2005. 1

[9] Z. Lei, Q. Bai, R. He, and S. Z. Li. Face shape recovery from a single image using cca mapping between tensor spaces. In *Proceedings of IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, pages 1–7, 2008. 2

[10] Z. Lei, S. Liao, and S. Z. Li. Coupled spectral regressoin for matching heterogeneous faces. In *Proceedings of IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, 2009. 2

[11] S. Z. Li. Heterogeneous face biometrics. In S. Z. Li, editor, *Encyclopedia of Biometrics*. Springer, 2009. 1

[12] 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*, 29(4):627–639, April 2007. 1, 2

[13] S. Liao, D. Yi, Z. Lei, R. Qin, and S. Z. Li. Heterogeneous face recognition from local structures of normalized appearance. In *Proceedings of IAPR/IEEE International Conference on Biometrics*, 2009. 1

[14] S. Liao, X. Zhu, Z. Lei, L. Zhang, and S. Z. Li. Learning multi-scale block local binary patterns for face recognition. In *Proceedings of IAPR/IEEE International Conference on Biometrics*, pages 828–837, 2007. 2

Table 3. Rank-1 recognition rates for Protocol I.

Methods	Exp.1	Exp.2	Exp.3	Exp.4	Exp.5	Exp.6
PCA-32	0.0550	0.0950	0.0200	0.0200	0.0100	0.0150
PCA-128	0.0600	0.1050	0.0200	0.0200	0.0100	0.0100
LDA-32	0.7700	0.8350	0.3300	0.4700	0.4600	0.4150
LDA-128	0.8050	0.8550	0.2400	0.3950	0.3300	0.3400

Table 4. Rank-1 recognition rates for Protocol II.

Methods	Exp.1	Exp.2	Exp.3	Exp.4	Exp.5	Exp.6
PCA-32	0.0667	0.0667	0.0351	0.0333	0.1228	0.0750
PCA-128	0.0667	0.0667	0.0175	0.0333	0.1228	0.0667
LDA-32	0.2750	0.3167	0.1228	0.2500	0.2632	0.2250
LDA-128	0.3083	0.4500	0.2982	0.2333	0.2632	0.2167

[15] D. Lin and X. Tang. Inter-modality face recognition. In *Proceedings of the European Conference on Computer Vision*, pages 13–26, 2006. 1

[16] J. Lu, K. N. Plataniotis, and A. N. Venetsanopoulos. Regularization studies on lda for face recognition. In *Proceedings of IEEE International Conference on Image Processing*, pages 63–66, 2004. 2

[17] MBGC 2008. http://face.nist.gov/mbgc/. 1

[18] P. J. Phillips, P. J. Flynn, W. T. Scruggs, K. W. Bowyer, J. Chang, K. Hoffman, J. Marques, J. Min, and W. J. Worek. Overview of the face recognition grand challenge. In *Proceedings of IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, pages 947–954, 2005. 2

[19] P. J. Phillips, H. Moon, S. A. Rizvi, and P. J. Rauss. The FERET evaluation methodology for face-recognition algorithms. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 22(10):1090–1104, 2000. 2

[20] T. Sim, S. Baker, and M. Bsat. The CMU pose, illumination, and expression database. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 25(12):1615–1618, 2003. 2

[21] X. Tan and B. Triggs. Enhanced local texture feature sets for face recognition under difficult lighting conditions. In *Proceedings of the IEEE International Workshop on Analysis and Modeling of Faces and Gestures*, 2007. 1

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

[23] 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. 4

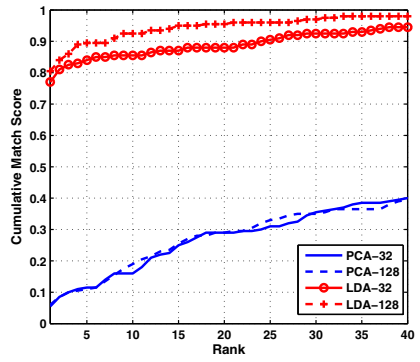
[24] R. Wang, J. Yang, D. Yi, and S. Z. Li. An analysis-by-synthesis method for heterogeneous face biometrics. In *Proceedings of IAPR/IEEE International Conference on Biometrics*, 2009. 2

[25] W. Yang, D. Yi, Z. Lei, J. Sang, and S. Z. Li. 2d-3d face matching using cca. In *Proc. IEEE International Conference on Automatic Face and Gesture Recognition*, 2008. 2

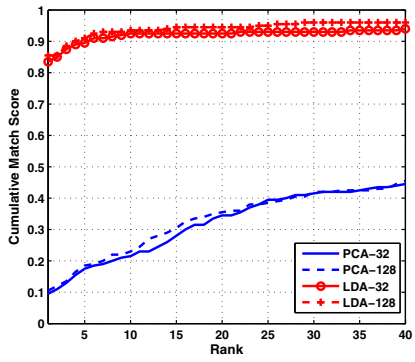
[26] D. Yi, S. Liao, Z. Lei, J. Sang, and S. Z. Li. Partial face matching between near infrared and visual images in mbgc portal challenge. In *Proceedings of IAPR/IEEE International Conference on Biometrics*, 2009. 1

[27] 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/IEEE International Conference on Biometrics*, pages 523–530, 2007. 1

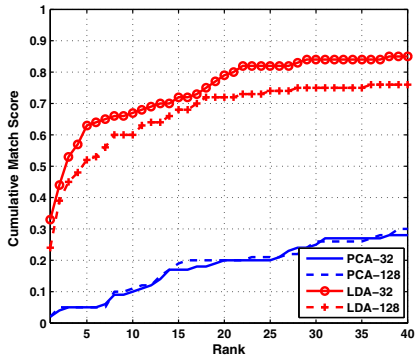
[28] D. Yi, R. Liu, R. Chu, R. Wang, D. Liu, and S. Z. Li. Outdoor face recognition using enhanced near infrared imaging. In *Proceedings of IAPR/IEEE International Conference on Biometrics*, pages 415–423, 2007. 3



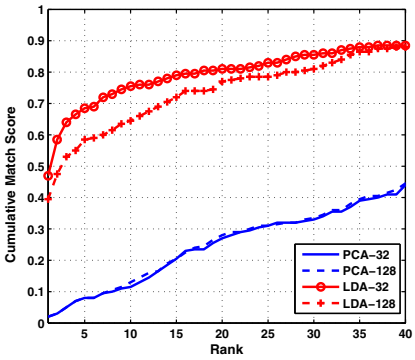
Exp. 1 (VIS-NIR)



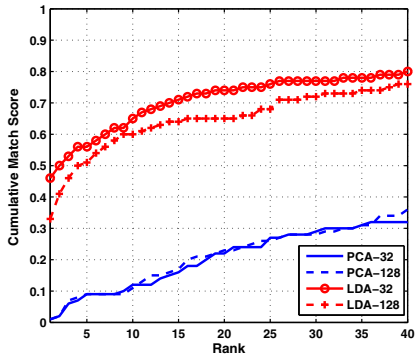
Exp. 2 (NIR-VIS)



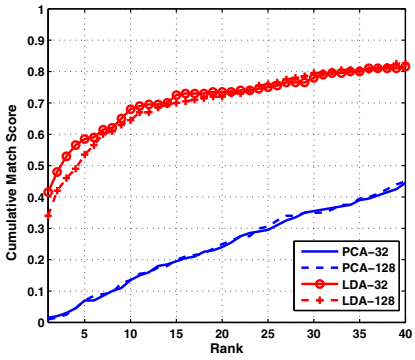
Exp. 3 (VIS-3D)



Exp. 4 (3D-VIS)

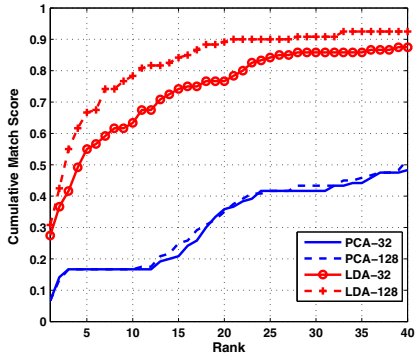


Exp. 5 (NIR-3D)

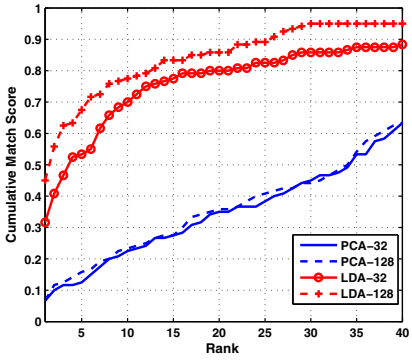


Exp. 6 (3D-NIR)

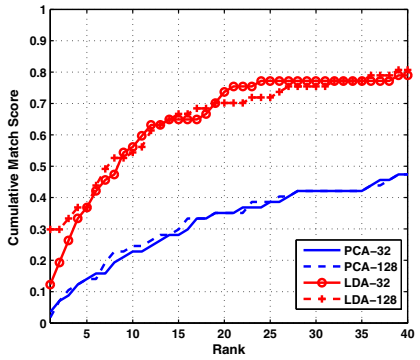
Figure 6. Cumulative match curves for protocol I.



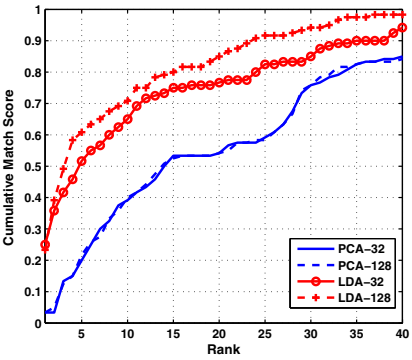
Exp. 1 (VIS-NIR)



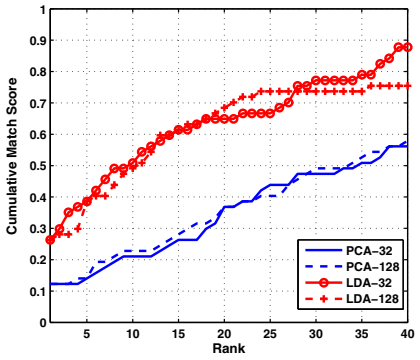
Exp. 2 (NIR-VIS)



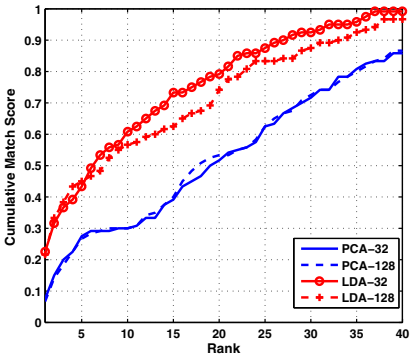
Exp. 3 (VIS-3D)



Exp. 4 (3D-VIS)



Exp. 5 (NIR-3D)



Exp. 6 (3D-NIR)

Figure 7. Cumulative match curves for protocol II.