Nighttime Face Recognition at Large Standoff: Cross-Distance and Cross-Spectral Matching

Dongoh Kang[a], Hu Han[b], Anil K. Jain[a,b,*], Seong-Whan Lee[a,*]

[a]*Department of Brain and Cognitive Engineering, Korea University, Anam-dong, Seongbuk-ku, Seoul 136-713, Republic of Korea*
[b]*Department of Computer Science and Engineering, Michigan State University, East Lansing, MI 48824, USA*

Abstract

Face recognition in surveillance systems is important for security applications, especially in nighttime scenarios when the subject is far away from the camera. However, due to the face image quality degradation caused by large camera standoff and low illuminance, nighttime face recognition at large standoff is challenging. In this paper, we report a system that is capable of collecting face images at large standoff in both daytime and nighttime, and present an augmented heterogeneous face recognition (AHFR) approach for cross-distance (e.g., 150m probe vs. 1m gallery) and cross-spectral (near-infrared probe vs. visible light gallery) face matching. We recover high-quality face images from degraded probe images by proposing an image restoration method based on Locally Linear Embedding (LLE). The restored face images are matched to the gallery by using a heterogeneous face matcher. Experimental results show that the proposed AHFR approach significantly outperforms the state-of-the-art methods for cross-spectral and cross-distance face matching.

Keywords:

Nighttime face recognition, heterogeneous face matching, cross-distance matching, cross-spectral matching, image restoration, K-means clustering, Locally Linear Embedding (LLE)

*Corresponding authors. Tel.: +82 2 3290 3197, E-mail: swlee@image.korea.ac.kr (S.-W. Lee), Tel.: +1 517 355 9282, E-mail: jain@cse.msu.edu (A. K. Jain)

1. Introduction

Surveillance cameras are being widely deployed in many countries worldwide for the purpose of traffic management, public safety control, and anti-terrorism. It is reported that there are over 4.2 million CCTV cameras in the United Kingdom, and on average, there is one camera for every 32 persons there [1]. There is also a tremendous growth in the number of surveillance cameras installed in major cities around the world [2]. Surveillance cameras have been found to be helpful to assist in law enforcement and criminal investigations in many high profile cases. As an example, in the case of Boston Marathon bombings on April 15, 2013, the surveillance videos with relative low quality (e.g., with motion blur), were found to be extremely useful in providing valuable clues to identify the suspects [3]. However, this still required a significant amount of manual processing of the video data. There is an urgent requirement for integrating automatic face recognition techniques into existing surveillance systems to improve the efficiency and accuracy of matching faces captured by surveillance cameras to mugshots and other face databases managed by law-enforcement and other government agencies.

Despite a significant progress in automatic face recognition, recognition of low-quality face images captured in unconstrained illumination and with arbitrary pose, typically encountered in surveillance videos, is a challenging problem [4]. While CCTV systems with visible light (VIS) cameras are able to capture relatively high-quality face images even at large standoff in daytime, VIS cameras do not work in adverse illumination conditions, e.g., in nighttime. However, many crimes occur in nighttime where the low illumination makes it difficult for most VIS image based CCTV systems to capture face images clearly (see Fig. 1 (a)). While the use of flash light is an option at nighttime for VIS cameras, it is not only expensive, but also not suitable in scenarios that require covert surveillance. By contrast, as shown in Fig. 1 (b), NIR cameras (with wavelength in the range 0.75–1.4 µm) can capture face images even at 150m standoff at nighttime that contain sufficient salient information for face recognition. Given that NIR face images are of a different sensing modality than VIS face images,
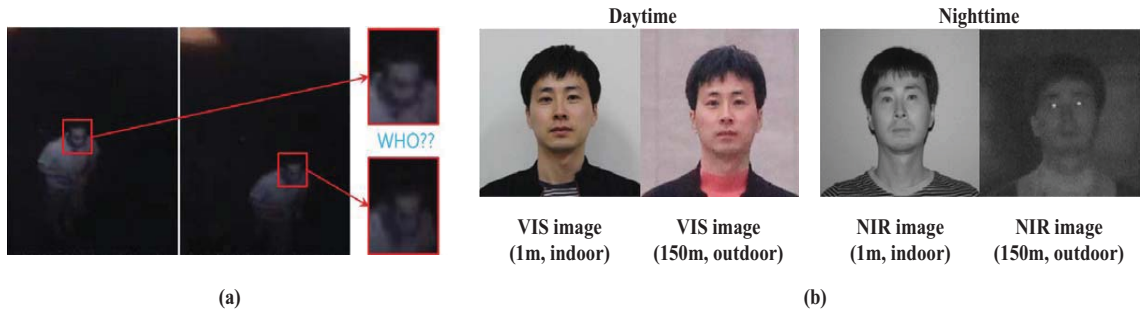
2

Figure 1: Visible light (VIS) vs. near-infrared (NIR) face images. (a) Visible spectral cameras in a surveillance system cannot capture high-quality face images suitable for recognition at nighttime even at a small standoff[1]. (b) NIR images that are captured at nighttime at a large standoff (e.g., 150m) still contain some discriminative facial details.

which typically reside in face databases maintained by law enforcement agencies, such as the mugshot and driver license databases, face recognition systems that are capable of performing cross-spectral (NIR probe to VIS gallery) and cross-distance (60m, 100m and 150m probe to 1m gallery) face matching are desirable in practical applications.

1.1. Problem Description and Contributions

In this paper, we address the cross-spectral (NIR vs. VIS) and cross-distance (60m, 100m, and 150m vs. 1m standoff) face matching problem for nighttime face recognition at large standoff. We have designed an image acquisition system for capturing face images in both daytime and nighttime at different standoffs (e.g., 1m, 60m 100m, and 150m). We integrated a telephoto lens with autofocus capability with a DSLR camera to capture VIS face images at large standoff in daytime. We further integrated an NIR illuminator into the system so that we can also capture NIR face images at large standoff in nighttime. With our image acquisition system, we have collected a Long Distance Heterogeneous Face (LDHF) database consisting of 1m indoor, and 60m, 100m, and 150m outdoor VIS and NIR images of 100 different subjects.

---

[1]http://www.suffolk.police.uk/newsandevents/newsstories/2013/february/
newmarketcctvappeal.aspx

Given the LDHF database, we have studied the problem of nighttime face recognition at a large standoff, and proposed an approach for cross-spectral and cross-distance face matching. Considering the image quality degradation (e.g., blur and noise) due to large standoff and low illuminance at nighttime, we first recover high-quality probe face images from the degraded probe face images by proposing an Locally Linear Embedding (LLE) based image restoration method. A heterogeneous face matcher is then used to match the restored face images to the 1m gallery face images.

The novelty of the proposed approach includes:

1. An image acquisition system designed to collect face images at large standoff in both daytime and nighttime, and a LDHF database that is publicly available to interested researchers[2].

2. A general framework for cross-spectral and cross-distance face matching that works for both NIR to VIS face matching in nighttime and VIS to VIS face matching in daytime.

3. A novel image restoration approach that significantly improves low-quality face images, and thus extends the applicability of existing heterogeneous face matchers to cross-spectral and cross-distance face matching.

The research reported in this paper is built upon our preliminary work [26]. The main differences between this paper and previously published paper are summarized as follows. (i) We propose an image restoration approach based on LLE to improve face image quality, which facilitates the application of existing face matchers to the scenarios of cross-spectral and cross-distance face matching. (ii) The proposed approach is compared with two Commercial-off-the-shelf (COTS) SDKs (FaceVACS [27] and PittPatt [28]) and two state-of-the-art heterogeneous face matchers (HFR [29] and LSNA [30]). (iii) The effectiveness of the proposed method is further evaluated in the scenario of face retrieval from two large gallery sets that are augmented with 10,000

---

[2]http://biolab.korea.ac.kr/database/

subjects from PCSO, and 2,508 Asian subjects extracted from PCSO, MORPH, CAS-PEAL, CASIA-FaceV5, CUHK, and FaceWarehouse databases, respectively.

1.2. Background

1.2.1. Cross-spectral Face Matching

Given that it is difficult to capture face images at nighttime using VIS cameras (see Fig. 1 (a)), different image acquisition methods have been developed to expand the capability of surveillance systems by collecting face images in different modalities than VIS, such as near-infrared (NIR), thermal, and SWIR images [5], [8], [9], [10], [11], [12], [13], [14]. Among these image acquisition methods, NIR images have been widely used in operational face recognition applications [15]. In particular, face recognition using NIR images has the following advantages [5], [6]: (i) NIR light is generally not visible to the human eyes and hence the surveillance operation can be somewhat covert; (ii) NIR images are less affected by ambient temperature, emotional and health conditions of the subjects compared to thermal images; (iii) NIR illuminators are cheaper compared to some other face imaging systems, such as thermal sensors; and (iv) NIR illumination can easily penetrate glasses, allowing it to capture more details around the eyes. These advantages of NIR imaging system have motivated a number of studies on face recognition using NIR images.

Existing approaches to NIR face recognition published in the literature can be grouped into three main categories: (i) Using NIR images [5], [7], [8], [9], [10], [16], [12] for both gallery and probe sets, (ii) image synthesis from NIR to VIS [18], [22], [23], and (iii) cross-spectral face recognition (or heterogeneous face recognition (HFR)) where the probe face image is NIR but the gallery face image is VIS. Approaches in the first category avoid the modality gap by enrolling NIR images in the gallery dataset; however, these approaches cannot be used in those applications where only VIS face images are enrolled and the NIR sensors are not available at the enrollment time. Approaches in the second category eliminate the modality gap by converting images with different modalities into the same modality, which makes it possible to perform face

Table 1: Representative face recognition methods using NIR face images reported in the literature

Problem domain	Publication	Face matching approach	Experimental protocol	
			Standoff	Environment
Intra-spectral Matching (NIR to NIR)	Zhao & Grigat 2005 [8]	DCT, SVM	∼1.5m	Indoor
	Huang et al. 2007 [5]	ELBP, Adaboost	∼1.5m	Indoor
	Shen et al. 2012 [9]	DBC, Adaboost	∼1.5m	Indoor
	Li et al. 2007 [10]	LBP, Adaboost	0.5m∼1m	Indoor
	Zhang et al. 2010 [16]	Gabor-DBC	0.8m∼1.2m	Indoor
	Bourlai & Cukic 2012 [12]	CSU face matcher	1.5m, 30m, 60m, 90m, 120m	Indoor outdoor
Image Synthesis (NIR vs. VIS)	Chen et al. 2009 [18]	LBP, LLE	∼1m	Indoor
	Wang et al. 2009 [22]	LoG	∼1.5m	Indoor
	Zhang et al. 2011 [23]	LBP, Sparse Representation	∼1.5m	Indoor
Heterogeneous Face Recognition (NIR vs. VIS)	Bourlai & Cukic 2012 [12]	CSU face matcher	1.5m, 30m, 60m, 90m, 120m	Indoor outdoor
	Yi et al. 2007 [24]	PCA-CCA,	∼1.5m	Indoor
	Liao et al. 2009 [30]	DoG, MB-LBP, Gentle Adaboost	∼1.5m	Indoor
	Klare & Jain 2010 [25]	HoG-LBP, LDA	0.8m∼1.2m	Indoor
	Klare & Jain 2013 [29]	MLBP, SIFT, LDA, Kernel	0.8m∼1.2m	Indoor
	Maeng et al. 2011 [11]	DoG-SIFT	1m, 60m	Indoor

recognition using traditional face matchers that have been developed for intra-spectral face matching. However, the synthesized images can only be called a pseudophoto due to their inferred content. Additionally, synthesis methods are often solving a more difficult problem than the recognition task. Approaches in the third category directly match NIR and VIS face images by using modality invariant face representations. Table 1 summarizes representative face recognition approaches using NIR face images reported in the literature.

Heterogeneous face recognition methods are particularly useful for nighttime face recognition where the probe images are NIR images, but the enrolled face gallery images are typically VIS photos. A number of studies related to HFR have been reported in recent years. Yi et al. [24] proposed a canonical correlation analysis (CCA) based method to learn the correlation between NIR and VIS images. Liao et al. [30] presented a Local Structure of Normalized Appearance (LSNA) method where common facial features in both NIR and VIS face images were extracted from Mult-scale Block LBP (MB-LBP) features using Gentle Adaboost. Klare and Jain [25] represented NIR and

<div style="text-align:center">(a)   (b)</div>

Figure 2: Nighttime NIR images at 60m standoff: (a) the NIR illuminator used in [11] introduces halo like light pattern around the head; (b) an improved NIR illuminator used in this paper leads to much better image quality compared to (a).

VIS face images using the histogram of oriented gradients (HOG) and LBP features, and then learnt discriminative subspaces with LDA to minimize the intra-subject differences due to the modality difference. Klare and Jain [29] also proposed a HFR framework where the NIR and VIS face images were represented in terms of nonliner similarities to a collection of prototype faces consisting of one image for each modality. Maeng et al. [11] normalized VIS and NIR images using DoG filter, and extracted SIFT features for cross-spectral and cross-distance matching. Bourlai and Cukic [12] studied both the NIR to VIS, and NIR to NIR face matching problems using a public CSU Face Identification Evaluation System [39]. Despite these methods proposed for NIR to VIS image matching, only a few studies have simultaneously considered the cross-spectral and cross-distance face matching problem [12], which is frequently encountered in surveillance scenarios.

1.2.2. Available Datasets and Their Limitations

A few databases have been collected for studying cross-spectral and cross-distance face recognition. Li et al. [35] collected both indoor VIS and NIR facial images, but they did not consider the scenario of face recognition at a large standoff. Maeng et al. [11] collected both indoor VIS and NIR facial images at two different standoffs (1m and 60m), but a noisy halo-like light pattern (see Fig. 2 (a)) appears in their NIR image

<div style="text-align:center">7</div>

Table 2: Summary of face image databases for cross-spectral and cross-distance matching

Publication		(#Images, #Subjects)	Standoff	Modality	Imaging environment	Available	Resolution
NFRAD [11]	Indoor	(600, 50)	1m, 60m	VIS & NIR	Daytime & Nighttime	No	3872×2592
SWIR [12]	DB1 Indoor	(NA, 50)	2m	VIS & SWIR	Controlled	No	NA
	DB2 Indoor	(NA, 50)	50m, 106m	VIS & SWIR	Semi-Controlled	No	NA
	DB3 Outdoor	(NA, 16)	60m 400m	VIS & SWIR	Uncontrolled	No	NA
NIR Mid-Range [12]	Indoor	(103, 103)*	1m	VIS	Controlled	No	NA
	Outdoor	(515, 103)*	30m, 60m, 90m, 120m	NIR	Nighttime	No	752×582
LDHF database (this study)	Indoor	(200, 100)	1m	VIS & NIR	Daytime & Nighttime	Yes	5184×3456
	Outdoor	(600, 100)	60m, 100m, 150m	VIS & NIR	Daytime & Nighttime	Yes	5184×3456

*These databases consist of video sequences instead of still images.

due to the NIR illuminator they used. Bourlai et al. [12] built a database with VIS gallery face images, and three types of probe images, including SWIR, MWIR, and NIR Mid-Range face images. The NIR Mid-Range dataset contains NIR face images captured at 30, 60, 90 and 120 meters in an outdoor environment. Table 2 summarizes the representative face databases for cross-distance and cross-spectral face matching reported in the literature.

Among these databases, only the NIR Mid-range dataset [12] simultaneously covers the scenarios of cross-spectral (NIR vs. VIS) and cross-distance (30m, 60m, 90m, and 120m) face matching. However, at large standoff, the database only contains NIR images captured outdoors at nighttime, and there are no VIS images captured outdoors in daytime. This makes it difficult to analyze the differences between NIR and VIS images in face recognition at large standoff. More importantly, the database in [12] is not available in the public domain. By contrast, the LDHF database we collected is available in the public domain and it includes both daytime VIS images and nighttime NIR images captured at 1m, 60m, 100m, and 150m standoff distances.

The remainder of this paper is organized as follows. In Section II we describe our face imaging system for collecting face images at a long distance in both daytime and nighttime. In section III we detail the proposed approach for cross-distance and cross-spectral face matching. We evaluate the proposed approach in Section IV, and conclude this work in Section V.
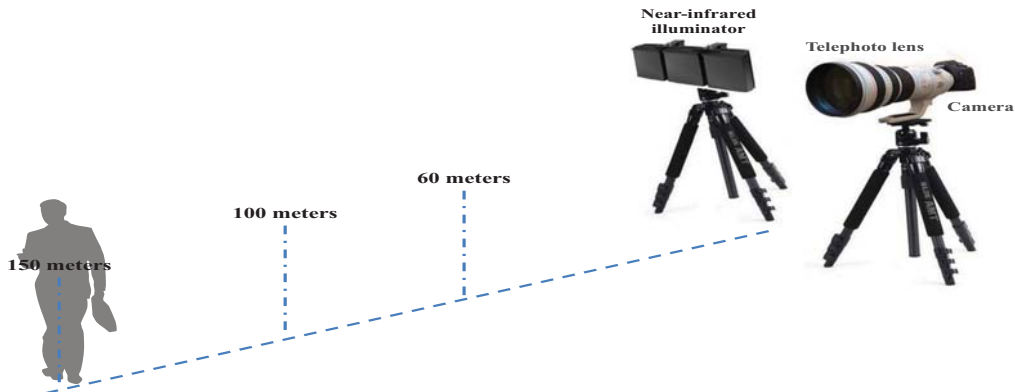
Figure 3: The proposed imaging system, designed for collecting face images at large standoff in both daytime and nighttime, consists of a high resolution DSLR camera (Canon 600D) with a telephoto lens (Canon EF 400mm F2.8L II IS USM) and a NIR illuminator (RayMax300).

2. LDHF Database Collection

2.1. LDHF Imaging System

To perform face recognition at a large standoff, we have designed an image acquisition system that is capable of capturing face images at long distances in both daytime and nighttime. As mentioned above, we capture NIR face images using an active NIR illuminator in nighttime. In practical applications, the NIR illuminators should meet the following requirements [15]: (1) Active NIR light should be nonintrusive to human eyes, (2) the direction of the NIR light emitted from the NIR illuminator should be fixed and consistent with the field of view of the camera, and (3) the reflected NIR light arriving at the camera sensor should override other light sources in the imaging environment. Given these requirements, we used the RayMax300 illuminator [38] which can emit constant NIR rays into a specific direction with a fixed wavelength of 850nm. The NIR rays from the RayMax300 illuminator are sufficiently powerful to reach up to 300 meters and is eye safe according to the product specification. The camera used in our setup is a Canon 600D DSLR with 18M effective pixels. In order to cover a large standoff during image acquisition, we used a Canon EF 400mm F2.8L II IS USM telephoto lens together with the DSLR. The Canon 600D DSLR has a built-in IR/UV cut-off filter in front of the CMOS sensor, so we removed the IR/UV filter in order to
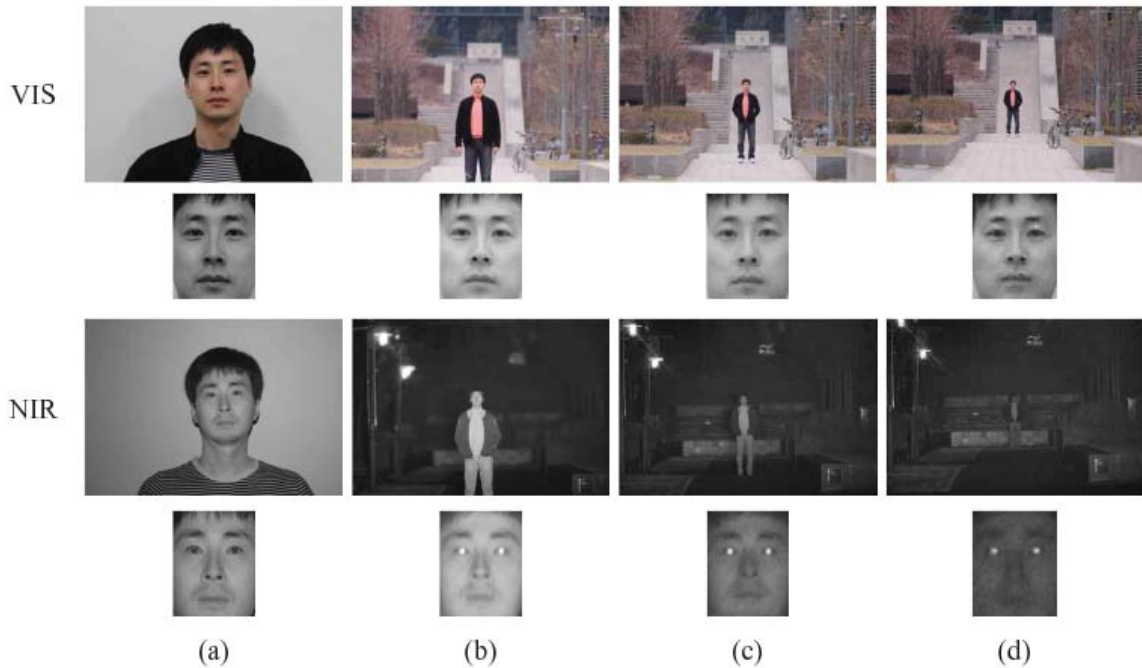
9

Figure 4: Example images in the LDHF database: VIS images (top row), cropped VIS face images with grey scale (second row), NIR images (third row), cropped NIR face images (bottom row) at (a) 1m, (b) 60m, (c) 100m, and (d) 150m.

capture the NIR light reflected by human faces. We also used the above image acquisition system to capture VIS face images at long distances in daytime without the NIR illuminator and without removing the IR/UV filter. Figure 3 shows the overview of the proposed face imaging system that was used to collect the LDHF database.

2.2. LDHF Database

Using the imaging system shown in Fig. 3, we have collected both VIS and NIR face images of 100 different subjects at 60m, 100m, and 150m standoff in an outdoor environment. For each subject, both frontal VIS and frontal NIR images were captured at each specified distance (standoff). In most face recognition applications, the enrolled gallery images are typically VIS images captured at a close standoff distance in a controlled environment. To replicate these scenarios, we also collected a 1m VIS face image indoors for each subject by using a lens with a short focal length (Canon

EF 50mm F1.8). Additionally, to cover the scenario of cross-distance NIR to NIR face matching, 1m NIR image was also captured for each subject indoors[4]. Therefore, as shown in Fig. 4, there are a total of 8 images were collected (four NIR images and four VIS images at 1m, 60m, 100m, and 150m, respectively) for each subject. The whole database was collected over a period of one month on the campus of Korea University, Seoul. The 100 subjects who participated in our study (70 males and 30 females) were students at Korea University with an age range of 20-30 years old. Some large standoff face images might look subtly different from each other because of the environmental conditions, such as fog and cloud. The image resolution is 5,184×3,456 pixels and the images are stored in both JPEG and RAW formats. The LDHF database is available to interested researchers via our website (`http://biolab.korea.ac.kr/database/`).

3. Cross-distance and Cross-spectral Face Matching

For nighttime face recognition, the gallery images are typically VIS images captured indoors, while the probe images are typically NIR images captured at various standoffs. As shown in Fig. 4, while VIS face images at different distances look quite similar, NIR face images show a significant quality degradation (e.g., photon or sensor noise) as the standoff distance increases. Additionally, NIR face images appear quite different from VIS face images because of the different sensing mechanism. To compensate for these problems, we first perform geometric and photometric normalization of face images to correct large pose (in-plane) and illumination variations. An image restoration method based on Locally Linear Embedding (LLE) is proposed to recover high-quality face images from the degraded images. Finally, we use a heterogeneous face matcher proposed in [29] to match the restored NIR probe images to the VIS gallery images. An overview of the proposed approach for cross-spectral and cross-distance face matching is shown in Fig. 5.

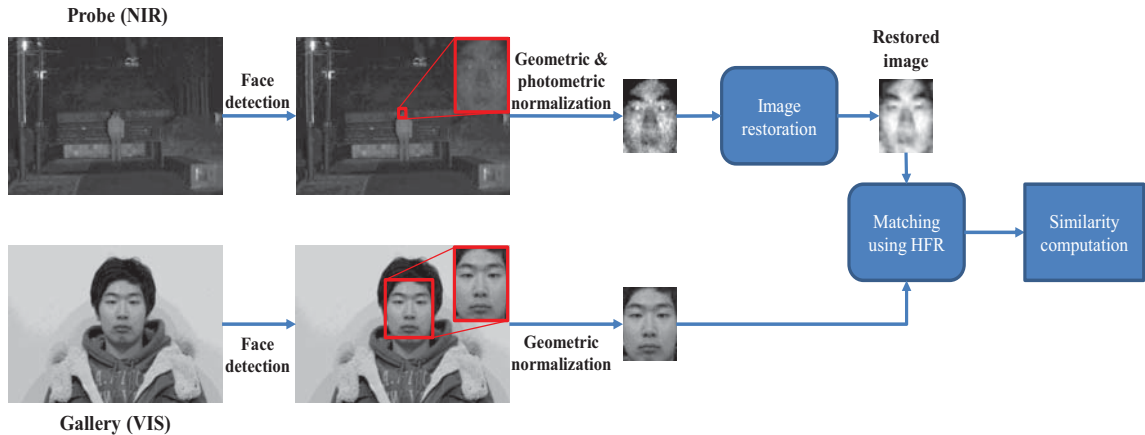[4]A less powerful NIR illuminator (24 IR LED) was used indoors.

Figure 5: Overview of the proposed algorithm (AHFR) for performing cross-spectral and cross-distance face recognition.

3.1. Geometric and Photometric Normalization

The purpose of geometric normalization is to reduce the effect of scale, rotation, and translation variations. The eye locations in both NIR and VIS face images are automatically detected using Cognitec's FaceVACS SDK 8.6.0 [27]. For face images where the eyes cannot be detected automatically (five VIS images at 150m, one NIR image at 100m, and 28 NIR images at 150m), we manually provided the eye locations. The face image is rotated and scaled so that the line connecting the two eyes is horizontal, and the interpupillary distance (IPD) is the same for all the images. Finally, all the face images are cropped to be of the same size (192×240). As shown in Fig. 6, to utilize the contextual information in the facial region [36], we scale the face images into the same size but with three different IPDs (126, 92, and 74 pixels). Specifically, we use a tight face cropping (IPD = 126 pixels) to remove the hair, forehead, ears, and jaw, and focus on the central facial region. We use a medium face cropping (IPD = 92 pixels) to include some portion of hair, forehead, and facial shape below the chin. Finally, we utilize a loose face cropping (IPD = 74 pixels) to include most of the hair, ears, jaw, and the whole facial shape. The proposed geometric normalization with multiple face croppings allows robustness to facial shape variations (e.g., different ratios between face
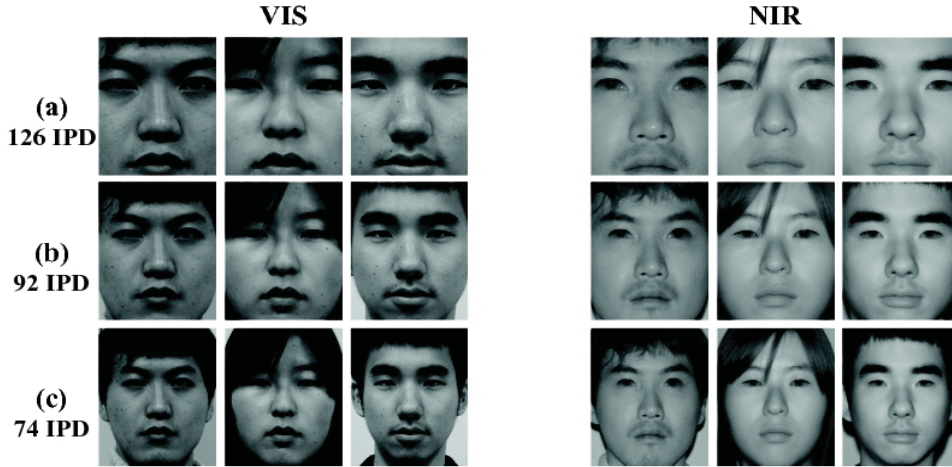
Figure 6: Three types of normalized faces with the same image size (192×240) but with different scales: (a) tight cropping (IPD = 126 pixels), (b) medium cropping (IPD = 92 pixels), and (c) loose cropping (IPD = 74 pixels).

height and width).

Photometric normalization has been found to be useful to improve the robustness of face recognition systems using VIS images [19], [20]. In this work, we propose a photometric normalization method which consists of noise filtering and contrast enhancement. We use 3×3 median filtering to suppress the high frequency noise while preserving the facial details. A holistic spatial contrast enhancement function [21] is then applied to the filtered image

$$\hat{\mathbf{I}} = (\mathbf{I} - m_i) \times v_e/v_i + m_e, \tag{1}$$

where $\mathbf{I}$ is the input face image; m_i and v_i are the mean and variance of input image $\mathbf{I}$, respectively; m_e and v_e are the desired mean and variance, respectively, which control the contrast of the output image $\hat{\mathbf{I}}$. In our experiments, $m_e = 100$ and $v_e = 80$. We do not apply photometric normalization to 1m VIS gallery face images captured indoors, because they are of sufficiently high quality and appropriate exposure.

3.2. Image Restoration

Maeng et al. [26] showed that the matching performance between 150m NIR and 1m VIS images was significantly lower than the matching performance between 60m NIR and 1m VIS images. Similar performance degradation was also observed in cross-distance VIS-to-VIS face image matching. We attribute the performance degradation to the image quality degradation arising from large standoff and low illuminance in nighttime. We propose a learning based image restoration method to recover a high-quality face image from a low-quality (large standoff) face image based on a training set consisting of pairs of low-quality face images and their corresponding high-quality face images. Specifically, we formulate the image restoration problem based on mani-fold learning theory, where the low-quality face images and their corresponding (mated) high-quality face images are considered to belong to low-quality and high-quality man-ifolds, respectively. Given a training set with both low-quality (e.g., 150m NIR) and the corresponding high-quality (e.g., 1m NIR) images, our objective is to learn a local linear mapping between the low-quality and high-quality manifolds, that can be used to recover the high-quality face images from any given low-quality face image. The proposed image restoration method consists of two main steps: 1) offline dictionary building based on the K-means clustering algorithm [32], and 2) online image restora-tion based on LLE. We describe these two steps in detail below.

3.2.1. Building the Dictionary

Let $\mathbf{S} = \{\{\mathbf{I}_i^L, \mathbf{I}_i^H\}, i = 1, 2, \cdots, m\}$ denote the training set, where $\mathbf{I}_i^L$ and $\mathbf{I}_i^H$ are the i-th normalized low-quality and high-quality images, respectively. For each pair of low-quality and high-quality images $\{\mathbf{I}_i^L, \mathbf{I}_i^H\}$, we randomly sample n pairs of corresponding patches $\{\mathbf{P}_{i,j}^L, \mathbf{P}_{i,j}^H\}, j = 1, 2, \cdots, n$, where each pair of patches comes from the same location in low-quality and high-quality face images (see Fig. 7). As a result, we have a total of $M = m \times n$ pairs of corresponding low-quality and high-quality patches, i.e., $\boldsymbol{\psi}^L = \{\mathbf{P}_{i,j}^L, i = 1, 2, \cdots, m; j = 1, 2, \cdots, n\}$ and $\boldsymbol{\psi}^H = \{\mathbf{P}_{i,j}^H, i = 1, 2, \cdots, m; j = 1, 2, \cdots, n\}$, respectively. For simplicity of notation, we rewrite $\boldsymbol{\psi}^L$ and $\boldsymbol{\psi}^H$ as $\boldsymbol{\psi}^L = \{\mathbf{P}_i^L, i =$
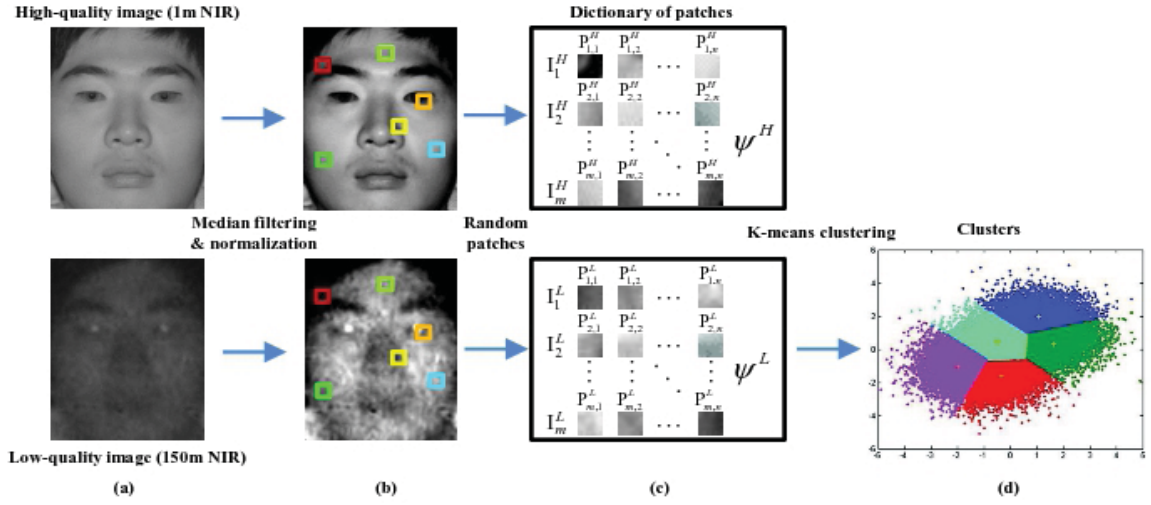
Figure 7: Overview of the dictionary building process for the proposed image restoration method: (a) input images, (b) normalized and filtered images, (c) dictionary, and (d) clustering.

$1, 2, \cdots, M\}$ and $\boldsymbol{\psi}^H = \{\mathbf{P}_i^H, i = 1, 2, \cdots, M\}$. The sets $\boldsymbol{\psi}^L$ and $\boldsymbol{\psi}^H$ constitute the dictionary that is used to learn a locally linear mapping between the low-quality and high-quality patches. In our experiments, we use pairs of low-quality face image and the corresponding 1m high-quality face image in the *same* spectral band to build $\boldsymbol{\psi}^L$ and $\boldsymbol{\psi}^H$; parameters are set as follows: $m = 90$, $n = 1100$, and two different patch sizes of 8×8 and 16×16.

In approaches where LLE has been used for super-resolution [17] and image synthesis [18], [40], each local patch in a test face image needs to be compared to all the M patches in $\boldsymbol{\psi}^L$ in order to determine a set of nearest-neighbor patches. In our experiments, this would require $M = 99,000$ comparisons between a patch in a test face image and $\boldsymbol{\psi}^L$ which can be slow in online testing stage. To reduce the computational cost, we propose a coarse-to-fine approach to determine the nearest-neighbor patches in $\boldsymbol{\psi}^L$. Specifically, we first group the low-quality patches in $\boldsymbol{\psi}^L$ into K clusters $\mathbf{C}_1^L, \mathbf{C}_2^L, \cdots, \mathbf{C}_K^L$ using K-means algorithm [32] (see Fig. 7), where the means of individual clusters are $\boldsymbol{\mu}_1^L, \boldsymbol{\mu}_2^L, \cdots, \boldsymbol{\mu}_K^L$. Patches in corresponding $\boldsymbol{\psi}^H$ are naturally partitioned into K clusters $\mathbf{C}_1^H, \mathbf{C}_2^H, \cdots, \mathbf{C}_K^H$ following the same procedure as for $\boldsymbol{\psi}^L$, where the cluster means are
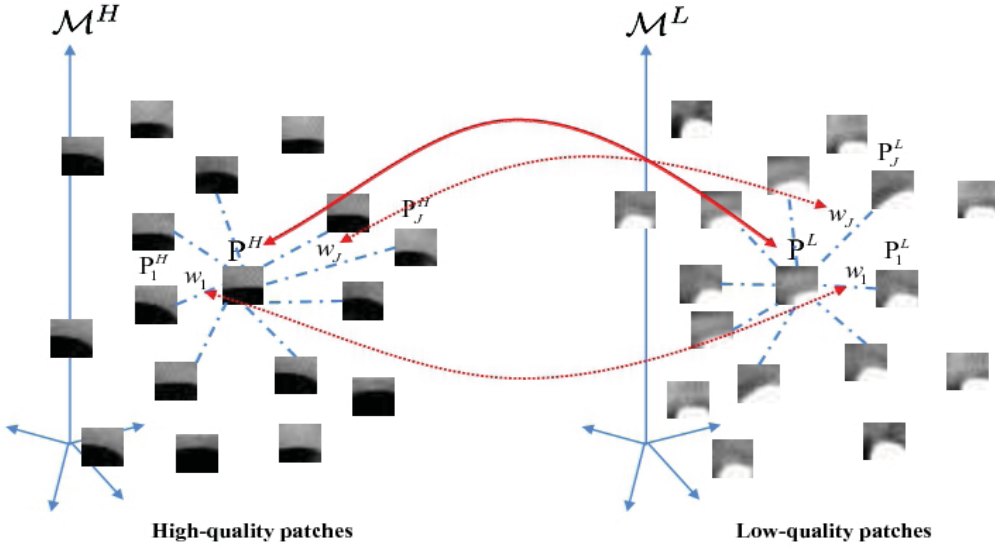
Figure 8: Preserving the local geometry between the low-quality manifold $\mathcal{M}^L$ and the high-quality manifold $\mathcal{M}^H$.

$\boldsymbol{\mu}_1^H, \boldsymbol{\mu}_2^H, \cdots, \boldsymbol{\mu}_K^H$. This way, the nearest-neighbor patches for a low-quality patch in a test face image can be determined by finding the nearest cluster in $\boldsymbol{\psi}^L$. In our experiments, we set $K = 512$ for the K-means clustering, so there are, on average, 200 (M/K) patches in each cluster. The number of comparisons needed to determine the nearest-neighbor patches for a low-quality patch in a test face image is reduced from $99{,}000$ (M) to 700 $(K + M/K)$ on average, which is over two orders of magnitude faster than the previous methods.

3.2.2. Restoring the Face Image

Following the manifold learning theory, two manifolds $\mathcal{M}^L$ and $\mathcal{M}^H$ can be learnt from $\boldsymbol{\psi}^L$ and $\boldsymbol{\psi}^H$, respectively, which have similar local geometry [17], [18], [40]. In other words, embeddings of a low-quality patch in $\mathcal{M}^L$ and the embedding of its corresponding high-quality patch in $\mathcal{M}^H$ have similar local geometry (see Fig. 8). Based on this observation, given a patch in a low-quality face image captured at a large standoff, its corresponding high-quality patch can be recovered by calculating its embedding in the high-quality manifold $\mathcal{M}^H$.
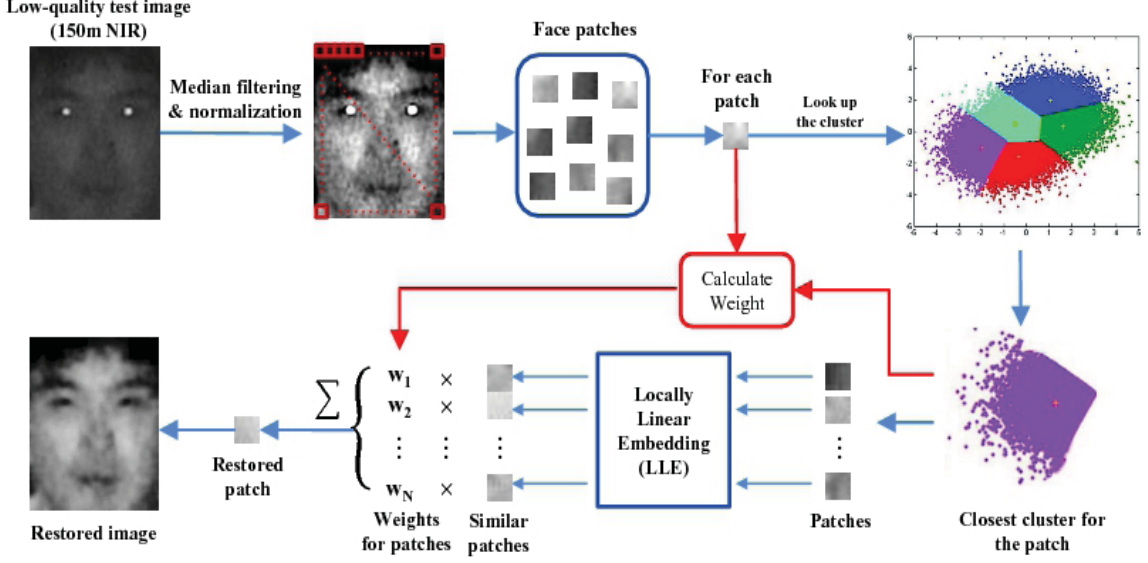
16

Figure 9: Proposed LLE based image restoration process.

Specifically, as shown in Fig. 9, given a low-quality face image, we first divide it into overlapping patches of the same size $w \times h$ which was used to build the dictionary. The overlap between two neighboring patches in the horizontal and vertical directions are $w_o = (w-1)$ and $h_o = (h-1)$ pixels, respectively. Thus, for a $W \times H$ low-quality face image, we have a total of $N = (W - w_o) \times (H - h_o)$ patches $\{\mathbf{P}_j^L, j = 1, 2, \cdots, N\}$. For a low-quality patch $\mathbf{P}_j^L$, we first find its closest cluster $\mathbf{C}_{\hat{k}}^L$ by comparing it with all the cluster means

$$\hat{\mathbf{k}} = \arg \min_k \|\mu_{\mathbf{k}}^{\mathbf{L}} - \mathbf{P}_j^L\|_{L^2}, \tag{2}$$

where $\| \cdot \|_{L^2}$ denotes the Euclidean distance. An embedding of patch $\mathbf{P}_j^L$ in the low-quality space $\mathcal{M}^L$ can then be calculated using LLE [31], which computes a set of weights $\{\omega_t, t = 1, 2, \cdots, T\}$ for the best linear reconstruction of $\mathbf{P}_j^L$ using its T neighboring patches $\mathcal{N}_j^L = \{\mathbf{P}_1^L, \mathbf{P}_2^L, \cdots, \mathbf{P}_T^L\}$ within cluster $\mathbf{C}_{\hat{k}^L}$

$$\varepsilon_j = \|\mathbf{P}_j^L - \sum_{\mathbf{P}_t^L \in \mathcal{N}_j^L} \omega_t \mathbf{P}_t^L\|_{L^2}. \tag{3}$$

Given the above weights $\{\omega_t, t = 1, 2, \cdots, T\}$ obtained by solving for T neighboring patches $\mathcal{N}_j^L$ in $\mathcal{M}^L$, the restoration of $\mathbf{P}_j^L$ can be computed from the corresponding T

17

neighboring patches in $\mathcal{M}^H$

$$\mathbf{P}_j^H = \sum_{\mathbf{P}_t^H \in \mathcal{N}_j^H} \omega_t \mathbf{P}_t^H, \tag{4}$$

where $\mathcal{N}_j^H = \{\mathbf{P}_1^H, \mathbf{P}_2^H, \cdots, \mathbf{P}_T^H\}$ consists of the patches in the $\mathcal{M}^H$ space that correspond to $\mathcal{N}_j^L$. However, the restoration using (4) can be sensitive to the specific T neighboring patches found in the $\mathcal{M}^L$ space. To reduce the sensitivity, we further consider the mean $(\boldsymbol{\mu}_{\hat{k}}^H)$ of cluster $\mathbf{C}_{\hat{k}}^H$ in the $\mathcal{M}^H$ space, and the restoration in (4) is modified to

$$\mathbf{P}_j^H = \frac{1}{2}\left(\mu_{\hat{k}}^H\right) + \frac{1}{2}\left(\sum_{\mathbf{P}_t^H \in \mathcal{N}_j^H} \omega_t \mathbf{P}_t^H\right). \tag{5}$$

After we perform restoration for all the N patches in an input low-quality face image, the overlapped regions of the restored patches are averaged to get the whole restored face image. In our experiments, we set the number of neighboring patches $T = 5$, and restore two high-quality images for each test image by using two different patch sizes, i.e., 8×8 and 16×16 for w×h.

3.3. Face Matching

After restoring high-quality face images via the proposed image restoration method, the "modality gap" between the VIS and NIR face images is significantly reduced. This makes it possible to utilize available heterogeneous face matchers developed for face matching at the same standoff (for both the VIS and NIR images) to perform face recognition. In our approach, we use the heterogeneous face recognition (HFR) system that was developed in [29] to match the restored face images at different standoffs to the 1m VIS gallery face images. As illustrated in Fig. 10, three different filters (Difference of Gaussian (DoG), Center-Surround Divisive Normalization (CSDN), and Gaussian) and two different descriptors (Scale Invariant Feature Transform (SIFT) [33] and Multiscale Local Binary Pattern (MLBP) [34]) were combined to extract six types of features from face images in HFR. All the features were extracted from overlapped patches and concatenated together to form a holistic representation of a face image. Random subspace LDA (RS-LDA) was then used to learn the discriminative projections that
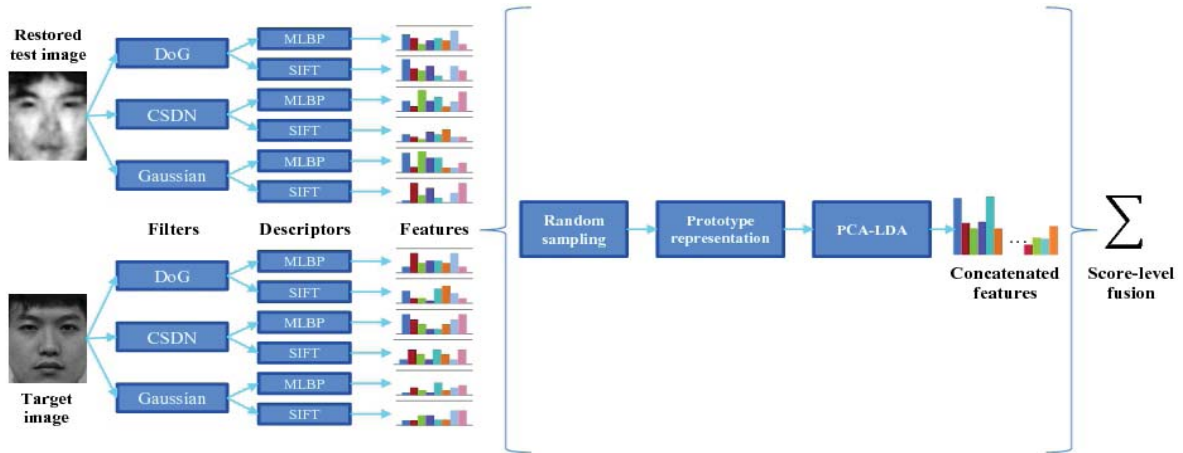
18

Figure 10: A block diagram of the heterogeneous face recognition (HFR) system [29] that is used to match the restored test face images at large standoff to target face images at 1m. The proposed augmented heterogeneous face recognition (AHFR) approach extends the applicability of HFR to cross-spectral and cross-distance face matching with a novel image restoration approach based on LLE.

minimizes the intra-person variations while maximizing the inter-person differences. Besides directly performing RS-LDA on the concatenated feature vectors, a prototype approach using kernel trick was further introduced to enhance the representation ability (prototype RS-LDA).

In our approach, we use HFR with prototype RS-LDA for cross-spectral and cross-distance face matching, and HFR with direct RS-LDA for intra-spectral and cross-distance face matching. Since we have cropped face images at three different scales (see Fig. 6), HFR is performed separately at individual scales. A score level fusion is used to calculate the final matching score.

4. Experimental Results

We evaluate the proposed approach using the LDHF database in three different scenarios: (i) cross-spectral and cross-distance face verification, (ii) intra-spectral and cross-distance face verification, and (iii) face identification with large gallery databases. The first scenario replicates the surveillance application in nighttime, where the query (probe) image is an NIR image while the target image is a VIS image. The second

scenario corresponds to surveillance applications in daytime or nighttime, where both the query and target images are of the same modality (either VIS images or NIR images). The last scenario commonly appears in forensic applications, where the mugshot of a suspect is used as the probe to find a match in large databases of mugshot photos. To compare the proposed approach with existing methods, we use two state-of-the-art methods (HFR [29] and LSNA [30]), and two COTS face matching systems (FaceVACS [27] and PittPatt [28]) as the baselines. Results of the two COTS matchers are denoted as "COTS-1" and "COTS-2" in order to anonymize the matchers. We also study the sensitivity of the parameters in the proposed image restoration approach to the face matching performance.

4.1. Experimental Settings

Due to the relatively small number of subjects (100) in the LDHF database, we perform image restoration using a 10-fold cross validation. In other words, in each image restoration experiment, 90 subjects are used to build the dictionary (ψ^L and ψ^H), and the remaining 10 subjects are used for testing. Since we obtain two different restorations of each low quality face image using two different patch sizes (8×8 and 16×16), the match scores using these two restored images are summed together to get the final similarity score.

We compare the performance of different methods by reporting the Receiver Operating Characteristic (ROC) curves for verification experiments in the first two scenarios, and Cumulative Matching Characteristic (CMC) curves for face identification experiments in the third scenario.

4.2. Cross-spectral and Cross-distance Face Matching

In this experiment, face verification experiments are performed by using 1m VIS images as target images, and NIR images at 60m, 100m and 150m as query images, respectively. We show the restored images using the proposed image restoration method in Fig. 11. Figure 11 (a) shows the input NIR images at different standoffs, where we can clearly see a degradation in image quality with an increase in standoff. Specifically,
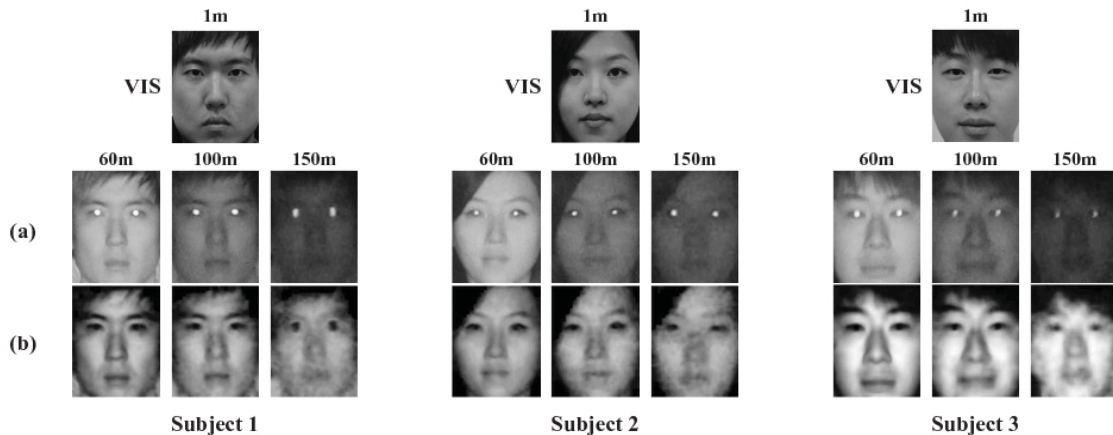
20

Figure 11: Examples of image restoration using the proposed approach for three different subjects: (a) input NIR face images and (b) restored NIR face images at three different standoffs. For each subject, we also show the 1m VIS gallery image.

most of the facial details are lost in the 150m NIR images. Figure 11 (b) shows the restored NIR images using the proposed image restoration approach. We can notice that while the original 60m NIR images look much better than the 100m NIR images, the restored 100m NIR images are quite consistent in terms of quality with the restored 60m NIR images. For the most challenging 150m NIR images, the proposed approach is still able to recover sufficient amount of facial details for improved face recognition.

The ROC curves of the proposed approach, two COTS matchers and two state-of-the-art heterogeneous face recognition methods are shown in Fig. 12. For 60m NIR images, the LSNA method and the two COTS matchers achieve similar performance. For example, at a FAR of 0.1%, the verification rate of LSNA and COTS-2 is ∼25%, and the verification rate of COTS-1 is ∼30%. On the other hand, the HFR method [29] with 50% verification rate at 0.1% FAR performs much better than LSNA and the two commercial matchers. However, the proposed approach achieves significantly higher recognition performance than HFR. For example, face verification rate of the proposed approach at 0.1% FAR is 75% compared with 50% verification rate for HFR.

For both 100m and 150m NIR images, while LSNA matcher still performs slight better than COTS-1, COTS-2 shows better performance than LSNA matcher. However,
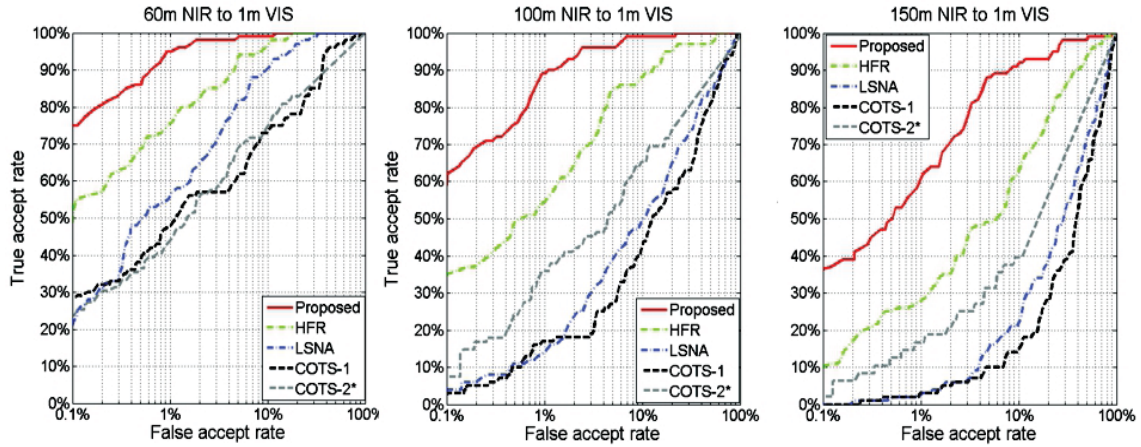
Figure 12: ROC curves for cross-spectral (NIR to VIS) and cross-distance (60m, 100m and 150m query images vs. 1m target images) matching.

A total of 1, 5, and 52 poor quality NIR face images at 60m, 100m and 150m, respectively, could not be enrolled by COTS-2 matcher.

this does not necessarily mean that COTS-2 is better in matching 100m and 150m NIR images to 1m VIS images, because some of the NIR images at 100m and 150m could not be enrolled (5 images at 100m and 52 images at 150m) by COTS-2. HFR still performs consistently better than the other three matchers at both 100m and 150m. However, the proposed approach achieves even more impressive improvement over HFR at 100m and 150m standoffs. For example, while HFR achieves 35% and 10% verification rates at 0.1% FAR for 100m and 150m NIR images, respectively, the proposed AHFR method achieves 60% and 37% verification rates.

From the above comparisons we can see the advantages of face matchers that are specifically designed for cross-spectral matching (HFR and LSNA) over the two commercial matchers mainly designed for VIS to VIS image matching. For example, LSNA is better than COTS-1 in most of the situations, and HFR is always better than both COTS-1 and COTS-2 in the conditions tested here. However, while LSNA and HFR were designed for handling cross-spectral matching problem, they are still limited to face matching with images captured at the same standoff. By contrast, the proposed AHFR approach is able to handle both cross-spectral and cross-distance scenarios.

22

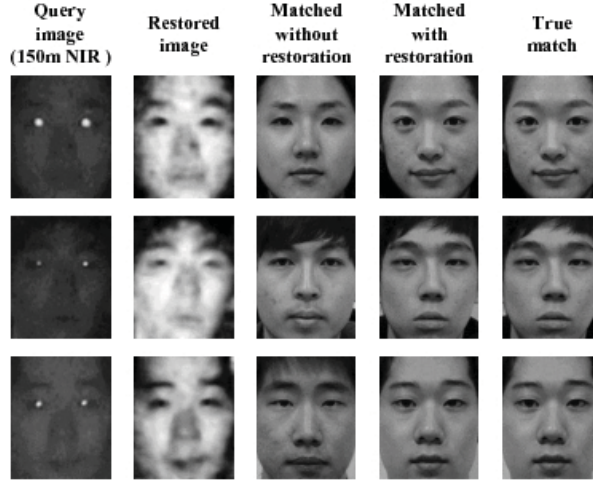| Query image (150m NIR)	Restored image	Matched without restoration	Matched with restoration	True match

Figure 13: Examples of face matching results at 0.1% False Accept Rate (FAR) by the proposed approach without and with using the restored images.

Figure 13 shows some examples of face matching results at 0.1% False Accept Rate (FAR) without and with using the restored images. Without using the proposed image restoration approach, the low-quality query images are matched to wrong target images (matched without restoration); however, after using the proposed image restoration approach, the query images are matched to the correct target images (matched with restoration). Figure 14 shows examples where the proposed approach leads to correct and incorrect matches. In Fig. 14 (b) we notice that some of the incorrect matches are due to the similarity in hair style. This is understandable because when many facial details are lost in low-quality NIR images captured at large standoff, the hair style indicated by the intensity changes becomes a dominant factor in face matching.

We also investigate the sensitivity of individual parameter values to the face matching performance. As shown in Fig. 15 (a), patch sizes of 8×8 and 16×16 used in image restoration are found to be better than the other patch sizes in matching 150m NIR images and 60m VIS images to 1m VIS images. Figure 15 (b) suggests that overlaps of 7 and 15 pixels are good choices for patch sizes of 8×8 and 16×16, respectively. Figures 15 (c) and (d) suggest that the matching performance is robust to the number of clusters, and the number of neighboring patches in LLE. In our experiments, we use
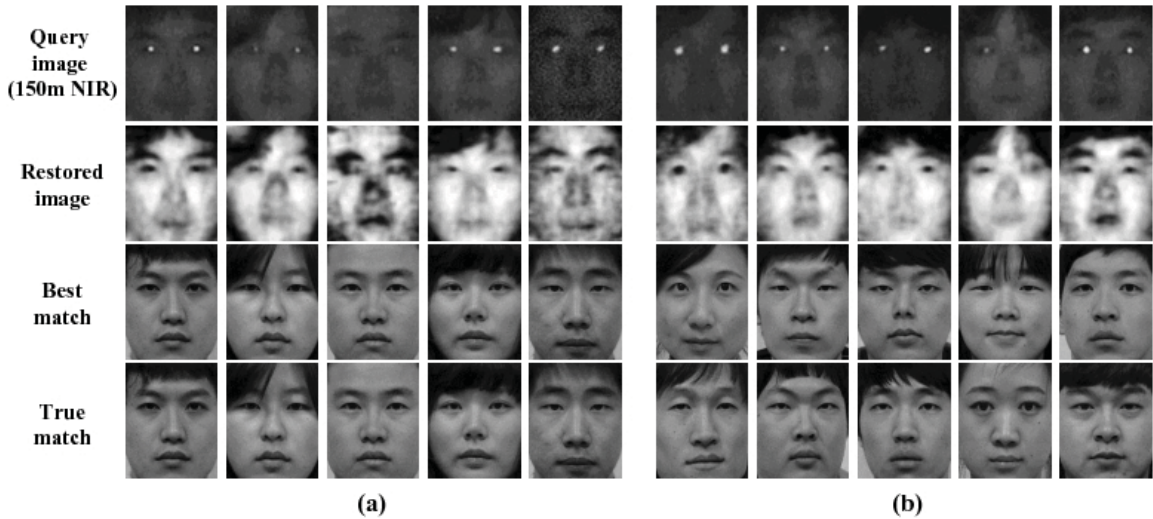
23

Figure 14: Examples of (a) correct matches, and (b) incorrect matches by the proposed approach at 0.1% FAR in cross-spectral and cross-distance matching.

512 clusters ($K = 512$), and 5 neighboring patches ($T = 5$) as suggested in [17], [18]. Table 3 shows that the proposed image restoration method with clustering is much more efficient than traditional restoration methods [17], [18] where LLE was used without clustering in the dictionary.

We also study the influence of the training set size by changing the percentage of face images used for training. Specifically, in our 10-fold test, 90% images are used for testing and 10% images are used for testing. In this experiment, while the testing set remains the same, we further reduce the training set size by using only 20%, 40%, 60%, and 80% images for training. Figure 16 shows the face verification results of both cross-spectral and intra-spectral matching at 60m, 100m and 150m with different percentages of training data. As shown in Fig. 16, although the percentage of training data varies from 20% to 90%, there are only minor differences in the performance of NIR-to-VIS and VIS-to-VIS matching. More importantly, the use of 90% images for training in a 10-fold cross-validation protocol does not lead to overfitting of the proposed cross-spectral and cross-distance face matching system. These results show the robustness of the proposed method against the training dataset size.
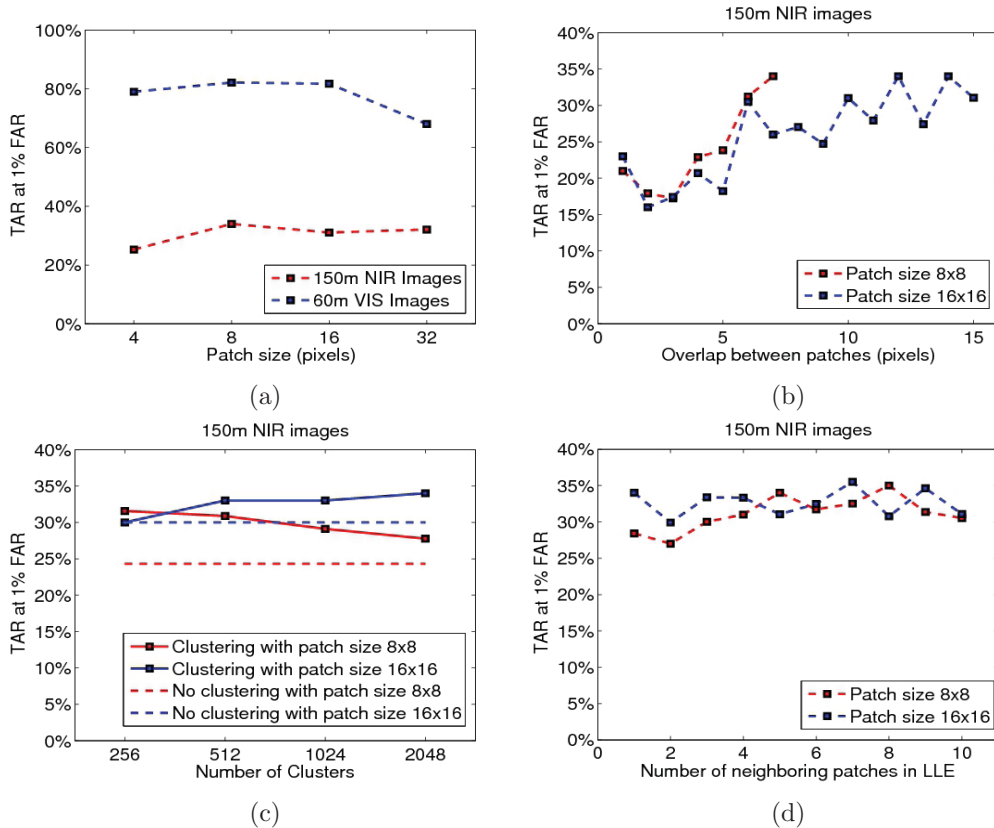
24

Figure 15: Sensitivity of algorithmic parameters on the cross-spectral cross-distance face matching performance: (a) patch size, (b) overlap between patches, (c) number of clusters in K-means, and (d) number of neighboring patches in LLE in the proposed image restoration approach.

4.3. Intra-spectral and Cross-distance Face Matching

In daytime face recognition at large standoff, the query images are typically of the same modality as the enrolled target images, i.e., both of them are VIS images. Hence, in this scenario, we are dealing with intra-spectral and cross-distance matching problem. Experiments in this scenario mainly evaluate the effectiveness of the proposed approach in the presence of difference in the standoff between the query and target face images.

Figure 17 (a) shows the ROC curves of various face recognition methods in matching 60m, 100m and 150m VIS query face images to 1m VIS target face images. At 60m, the proposed AHFR approach achieves significantly higher performance (91%) than HFR (75%), COTS-2 (69%), and LSNA (55%), but slightly lower performance than
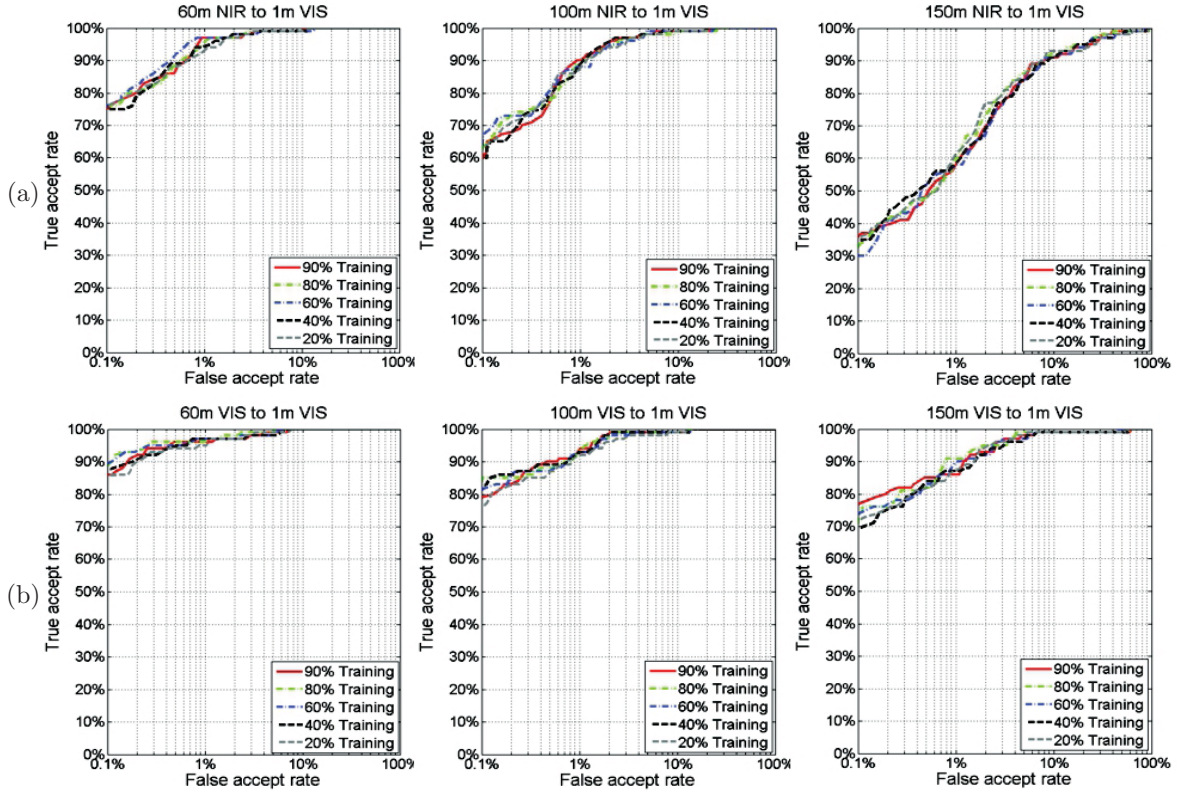
Figure 16: Influence of different percentages (20%, 40%, 60%, 80%, and 90%) of training data for (a) cross-spectral (NIR to VIS) and (b) intra-spectral (VIS to VIS) matching at 60m, 100m, and 150m standoff.

the commercial matcher COTS-1 (96%) at 0.1% FAR. The main reason for the better performance of COTS-1 is probably because of the fact that 60m VIS face images captured using our image acquisition system are still of quite high quality (see Fig. 4). However, for larger standoff, namely 100m and 150m, blur and noise are observed in the face images, and COTS matchers no longer outperform the designed cross-distance face matching method. For example, as shown in Fig. 17 (a), when the standoff distance increases from 60m to 100m, the performance of COTS-1 matcher at 0.1% FAR drops from 96% to 67%, but the performance of the proposed AHFR approach only drops from 91% to 83%, which shows better robustness against standoff. At 150m, the proposed AHFR approach still achieves 75% verification rate at 0.1% FAR, which is much higher than the best state-of-the-art method (55% verification by COTS-2). The effectiveness

Table 3: Average computation time (in seconds) for image restoration with and without K-means clustering on a laptop with 2.5G dual-core Intel Core i5 processor and 16G RAM.

	Average time (s) per image	
Patch size	No clustering	Clustering with 512 clusters
8×8	528.29	5.00
16×16	1301.37	10.17

of the proposed AHFR approach in VIS to VIS face matching across different standoffs also encouraged us to evaluate the proposed approach in NIR to NIR face matching across different standoffs. Figure 17 (b) shows that the proposed AHFR approach is also more robust than existing methods in NIR to NIR face matching across different standoffs. The above experimental results show that the proposed AHFR approach is more effective than the state-of-the-art methods in intra-spectral and cross-distance face recognition (see Figs. 18 (a) and (b)). Examples of correct and incorrect matches by the proposed approach for intra-spectral cross-distance matching are shown in Fig. 19.

It is generally true that in most face recognition systems, only VIS images of individual subjects are enrolled in the target set. However, with the development of NIR imaging devices, it is also possible to enroll the NIR images of individual subjects in the target set. In this situation, it is natural to utilize both VIS and NIR target images to improve the face matching performance. The availability of both 1m VIS and 1m NIR images in our LDHF database makes it possible for us to investigate this problem. So, in our experiment, a query NIR image is matched to both the 1m VIS and 1m NIR target images using the proposed approach, and the resulting two similarity scores are summed up to get the final match score for the query and a target subject. We show the ROC curves with 150m NIR images as the query images in Fig. 18 (c). We notice that a simple score level fusion of the two modalities of target images is able to further improve the nighttime face recognition performance (see Fig. 20).
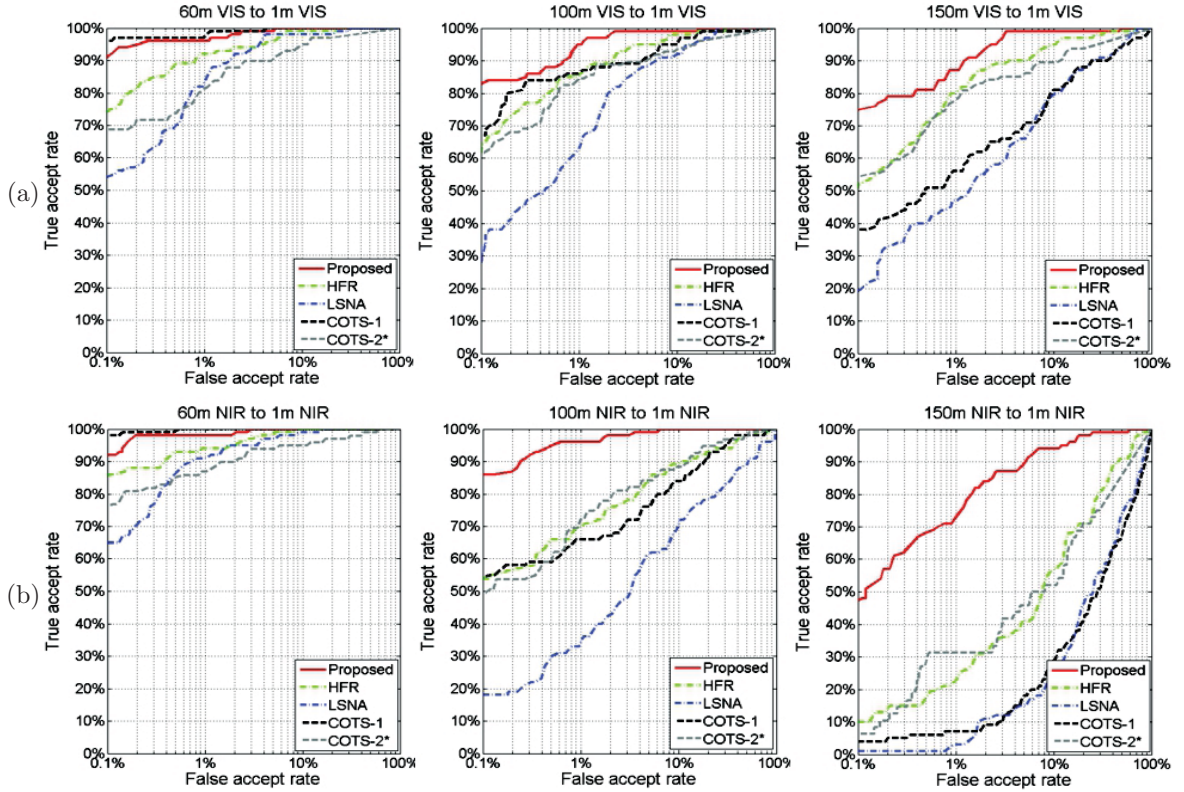
Figure 17: ROC curves for intra-spectral (VIS to VIS or NIR to NIR) and cross-distance (60m, 100m and 150m query images vs. 1m target images) matching.

A total of 1, 5, and 52 images, respectively, could not be enrolled by COTS-2 matcher for 60m, 100m, and 150m NIR images due to poor quality; similarly, a total of 1, 3, and 6 images, respectively, could not be enrolled by COTS-2 matcher for 60m, 100m, and 150m VIS images.

4.4. Identification with a Large Gallery

In forensic applications, the NIR image of a suspect captured at nighttime is likely to be used to perform retrieval from a large mugshot database to determine the identity of the suspect. To better replicate this scenario, we augment the gallery set (target set used in Sections 4.2 and 4.3) of 100 subjects with 10,000 additional mugshots (one image per subject) from the Pinellas County Sheriff's Office (PCSO) database[5]. The 10,000 background images from PCSO are mugshots with slight pose, expression, and illumination variations (see Fig. 21 (a)). Face identification (closed set) is then performed
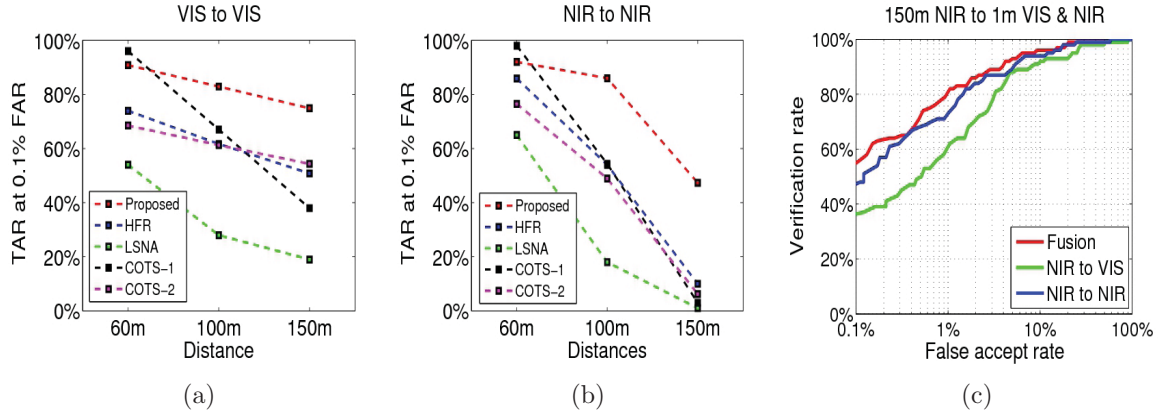
28

Figure 18: Robustness of the proposed approach in (a) cross-distance VIS to VIS face matching, (b) cross-distance NIR to NIR face matching, and (c) fusion of VIS and NIR target images in matching 150m NIR images to 1m VIS and 1m NIR images.

using this enlarged VIS gallery set.

Figure 22 (a) shows the CMC curves of face identification using 60m, 100m and 150m NIR images in an operational rank range, e.g., rank-1 to rank-200 [37]. We can see that HFR [29] achieves better performance than the other three state-of-the-art methods (COTS-1, COTS-2 and LSNA). For examples, the rank-1 accuracies of HFR at 60m, 100m and 150m are 65%, 48%, and 15%, respectively. Compared with HFR, the proposed AHFR approach achieves much higher performance with the rank-1 accuracies of 82%, 69%, and 28% for 60m, 100m and 150m NIR images, respectively. At rank 200, the proposed AHFR approach achieves 98%, 94%, and 76% accuracies for 60m, 100m and 150m NIR images, respectively, which are better than HFR accuracies (96%, 88%, and 54% accuracies for 60m, 100m and 150m NIR images, respectively). Similar experiments were also performed to match VIS probe images to the enlarged VIS image gallery set. As shown in Fig. 22 (b), the proposed AHFR approach achieves comparable performance to COTS-1 for 60m VIS probe images, but slight better performance than the state-of-the-art methods for 100m and 150m VIS probe images. Examples of correct

---

[5]PCSO is a mugshot database that is collected by the Pinellas County Sheriff's Office (PCSO), Florida.
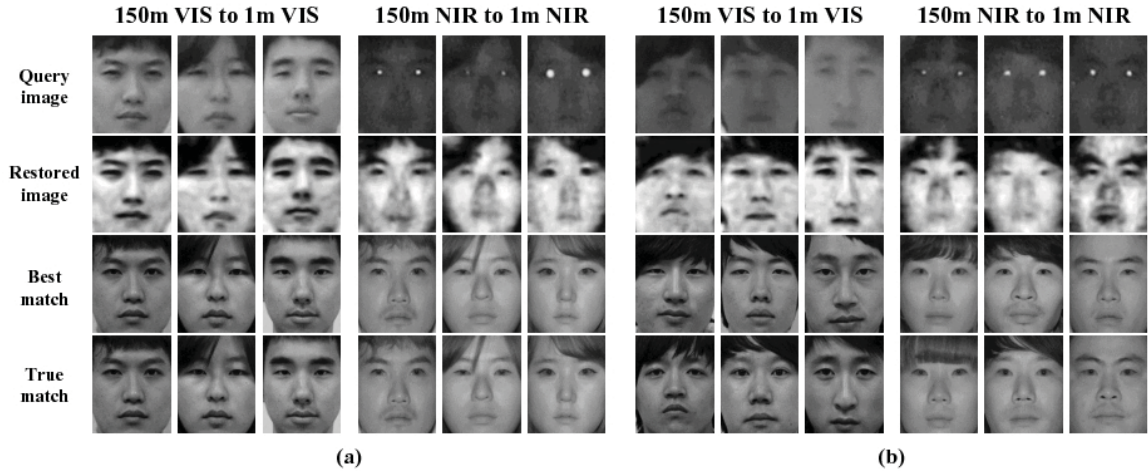
Figure 19: Examples of (a) correct matches, and (b) incorrect matches by the proposed approach at 0.1% FAR in intra-spectral and cross-distance matching.

and incorrect rank-1 matches by the proposed approach are shown in Fig. 23.

To determine the effectiveness of our approach using a background database with Asian subjects (similar ethnicity as the faces in our LDHF database), we collected an Asian background set by identifying the Asian subjects in the following databases: PCSO, MORPH [41], CAS-PEAL [42], CASIA-FaceV5 [43], CUHK [44], and FaceWarehouse [45] (see Fig. 21 (b)). This enabled us to build an Asian background set with 2,508 subjects (one image per subject) to populate the gallery. To our knowledge, there is no large public domain Asian face database available. Figure 24 shows the face identification performance of the proposed AHFR approach for cross-spectral and intra-spectral face matching at different distances. The performance of two COTS matchers and two state-of-the-art heterogeneous face matching methods (HFR and LSNA) are also provided for comparison. Figure 24 shows that the proposed AHFR approach performs significantly better than other baseline methods for NIR to VIS matching at different distances using the background dataset with Asian subjects. For VIS to VIS matching, the proposed approach achieves comparable performance to state-of-the-art matchers at 60m, but slightly better performance than state-of-the-art matchers at 100m and 150m. These observations with an Asian background set are consistent with those that
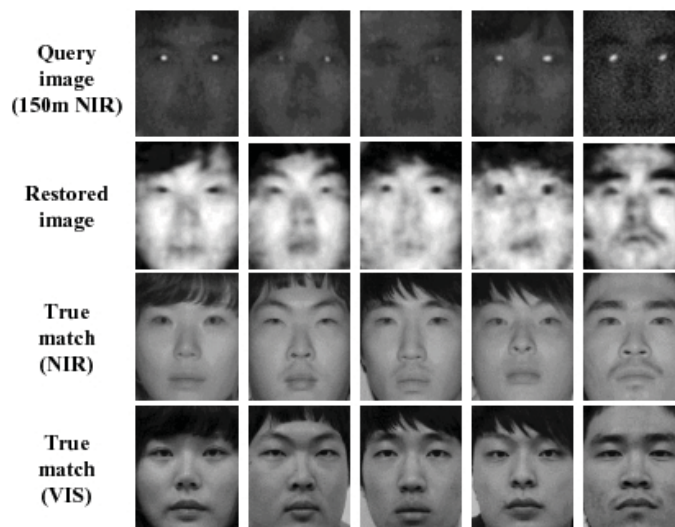
30

Figure 20: Examples of 150m NIR image (top row) that could not be correctly matched when compared only to 1m VIS target, but could be correctly matched when compared to both 1m VIS (third row) and 1m NIR target (bottom row) images at FAR = 0.1%.

we observed when the PCSO dataset was used as the background set.

5. Conclusions and Future Work

We have studied the problem of cross-spectral (NIR vs. VIS) and cross-distance (60m, 100m and 150m vs. 1m) face matching problem to address the application of face recognition in nighttime operations. We have designed an image acquisition system that is able to collect face images at large standoff for both VIS images in daytime and NIR images at nighttime. With this imaging system, we have collected a face database, called Long Distance Heterogeneous Face (LDHF) database that has been made publicly available. Given the LDHF database, we studied the nighttime face recognition problem at large standoff, and proposed the AHFR approach for cross-spectral cross-distance matching.

The proposed AHFR approach was evaluated under three different scenarios: (i) cross-spectral and cross-distance face matching, (ii) intra-spectral and cross-distance face matching, and (iii) identification with a large gallery. Experimental results show

31

Figure 21: Examples of background VIS face images. (a) Mugshots with pose and expression variations in the PCSO mugshot database, and (b) Asian subjects taken from PCSO, MORPH, CAS-PEAL, CASIA-FaceV5, CUHK and FaceWarehouse databases.

that the proposed AHFR approach achieves significantly better performance under all the three scenarios compared to four baselines: two COTS matchers (designed for VIS to VIS matching) and two state-of-the-art heterogeneous face matching methods, HFR [29] and LSNA [30], designed for NIR to VIS matching. One of the main contributions of our work that is responsible for improved cross-spectral and cross-distance face recognition performance of the proposed method is the learning based face image restoration method.

Our future work will include: (i) improving the NIR face image acquisition system and enlarging the database size (e.g., 200 subjects with multiple sessions), (ii) introducing some degree of pose, expression, and occlusion variations during face image capture, and (iii) finding a face representation for further improving the cross-spectral and cross-distance face recognition accuracy.
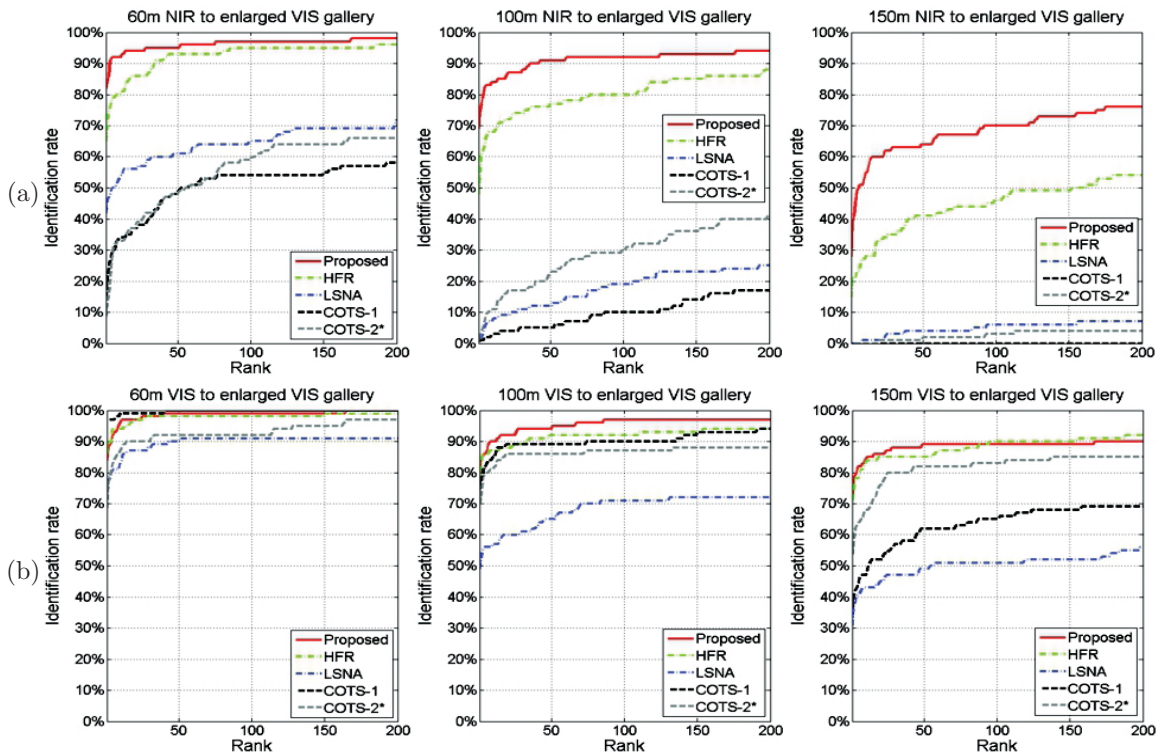
Acknowledgement

Figure 22: CMC curves for cross-spectral (NIR to VIS) and intra-spectral (VIS to VIS), and cross-distance (60m, 100m, and 150m probes to 1m VIS gallery) face identification using additional 10,000 background face images from PCSO.

Some face images could not be enrolled by COTS-2 matcher. For example, 1, 5, and 52 images could not be enrolled by COTS-2 matcher for 60m, 100m, and 150m NIR images, respectively; and 1, 3, and 6 images could not be enrolled by COTS-2 matcher for 60m, 100m, and 150m VIS images, respectively. Among the gallery images, 8 mugshots from PCSO could not be enrolled by COTS-2 matcher.

[1] Big Brother is DEFINITELY watching you: Shocking study reveals UK has one CCTV for every 32 people (2011), [Online] http://www.dailymail.co.uk/news/article-1362493/One-CCTV-camera-32-people-Big-Brother-Britain.html.

[2] C. Norris, A Review of the Increased Use of CCTV and Video Surveillance for Crime Prevention Purposes in Europe, European Parliament, 2009. [Online] http://www.statewatch.org/news/2009/apr/ep-study-norris-cctv-video-surveillance.pdf.
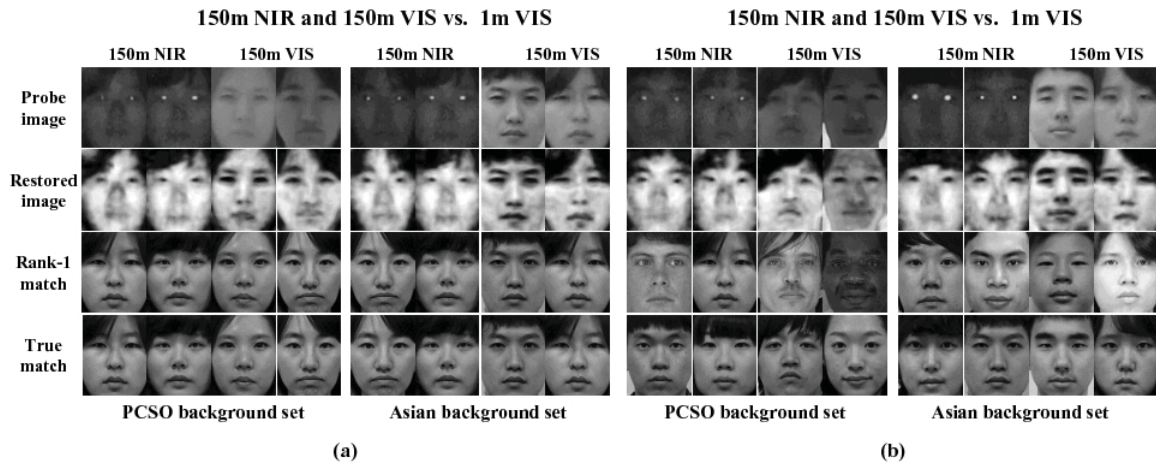
Figure 23: Examples of (a) correct and (b) incorrect rank-1 matches in face identification using 150m NIR probe face images and 1m VIS gallery images with two different background sets (a PCSO dataset with 10,000 images, and an Asian dataset with 2,508 images).

[3] J. C. Klontz, A. K. Jain, A Case Study on Unconstrained Facial Recognition using the Boston Marathon Bombings Suspects, MSU Technical Report, 2013. [Online] `http://www.cse.msu.edu/biometrics/Publications/Face/KlontzJain_CaseStudyUnconstrainedFacialRecognition_BostonMarathonBombimgSuspects.pdf`.

[4] P. J. Phillips, P. J. Flynn, J. R. Beveridge, W. T. Scruggs, A. J. O'Toole, D. Bolme, K. W. Bowyer, B. A. Draper, G. H. Givens, Y. M. Lui, H. Sahibzada, J. A. Scallan, W. Samual, Overview of the Multiple Biometrics Grand Challenge, in Proc. ICB, 2009, pp. 705-714.

[5] D. Huang, Y. Wang, Y. Wang, A Robust Method for Near Infrared Face Recognition Based on Extended Local Binary Pattern, in Proc. ISVC, 2007, pp. 437-446.

[6] X. Zou, J. Kittler, K. Messer, Illumination Invariant Face Recognition: A Survey, in Proc. IEEE BTAS, 2007.

[7] Z. Pan, G. E. Healey, M. Prasad, B. J. Tromberg, Face Recognition in Hyperspectral Images, IEEE Trans. Pattern Anal. Mach. Intell. 25 (2003) pp. 1552-1560.
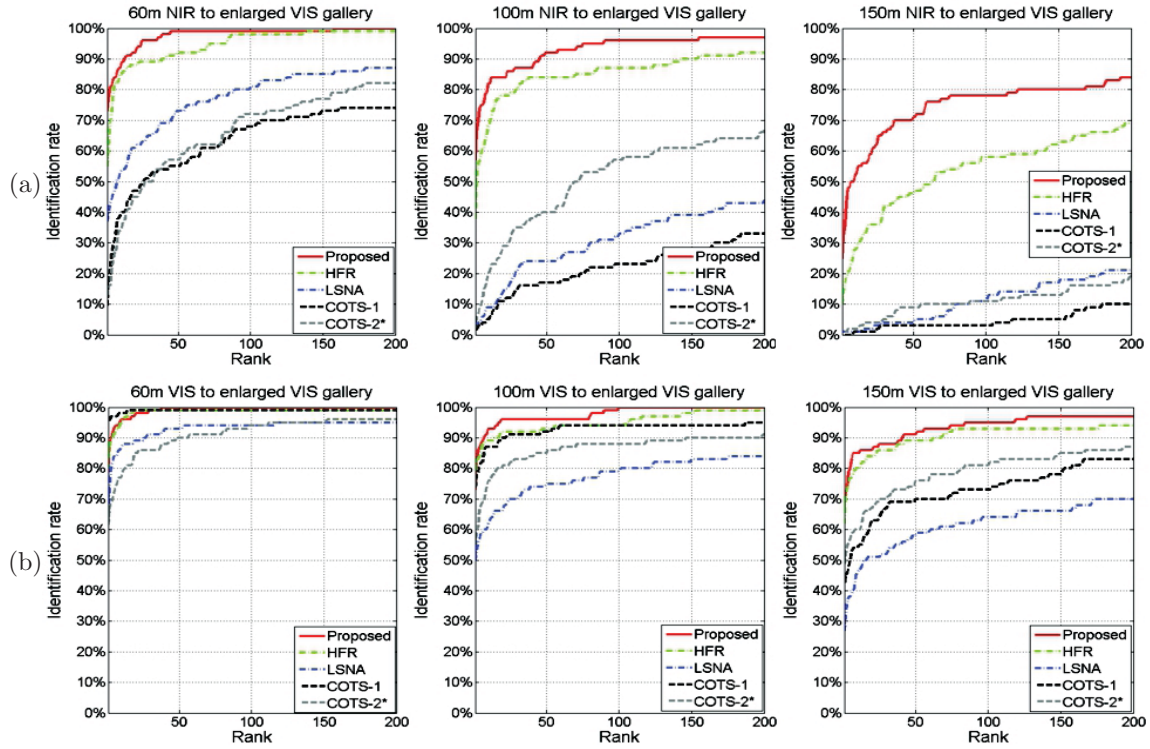
Figure 24: CMC curves for (a) cross-spectral (NIR to VIS) and (b) intra-spectral (VIS to VIS) face matching at 60m, 100m, and 150m with a background set of Asian subjects (assembled from PCSO, MORPH, CAS-PEAL, CASIA-FaceV5, CUHK, and FaceWarehouse databases).

Some of the low-quality face images at large standoff could not be enrolled by COTS-2 matcher. A total of 1, 5, and 52 images, respectively, could not be enrolled by COTS-2 matcher for 60m, 100m, and 150m NIR images; and 1, 3, and 6 images, respectively, could not be enrolled by COTS-2 matcher for 60m, 100m, and 150m VIS images. Among the gallery images, 4 mugshots from PCSO, MORPH, CAS-PEAL, CASIA-FaceV5, CUHK, and FaceWarehouse could not be enrolled by COTS-2 matcher.

[8] S. Zhao, R.-R. Grigat, An Automatic Face Recognition System in the Near Infrared Spectrum, in Proc. MLDM, 2005, pp. 437-444.

[9] L. Shen, J. He, S. Wu, S. Zheng, Face Recognition from Visible and Near-Infrared Images using Boosted Directional Binary Code, in Proc. ICIC, 2012, pp. 404-411.

[10] S. Z. Li, R. Chu, S. Liao, L. Zhang, Illumination Invariant Face Recognition Using Near-Infrared Images, IEEE Trans. Pattern Anal. Mach. Intell. 29 (2007) pp. 627-639.

[11] H. Maeng, H.-C. Choi, U. Park, S.-W. Lee, A. K. Jain, NFRAD: Near-Infrared Face Recognition at a Distance, in Proc. IJCB, 2011, pp. 1-7.

[12] T. Bourlai, B. Cukic, Multi-Spectral Face Recognition: Identification of People in Difficult Environments, in Proc. ISI, 2012, pp. 196-201.

[13] F. Nicolo, N. A. Schmid, Long Range Cross-Spectral Face Recognition: Matching SWIR Against Visible Light Images, IEEE Trans. Inf. Forensics Security 7 (2012) pp. 1717-1726.

[14] Tactical Imager for Night/Day Extended-Range Surveillance WVHTCF, Nov. 10, 2011 [Online]. Available: `http://www.wvhtf.org/programs/advancedtech/active_swir_imaging_for_long_range_nighttime_surveillance_and_face_recognition.php`.

[15] S. Z. Li, A. K. Jain (eds.), Handbook of Face Recognition, Springer, 2011.

[16] B. Zhang, L. Zhang, D. Zhang, L. Shen, Directional Binary Code with Application to PolyU Near-Infrared Face Database, Pattern Recogn. Lett. 31 (2010) pp. 2337-2344.

[17] H. Chang, D.-Y. Yeung, Y. Xiong, Super-Resolution Through Neighbor Embedding, in Proc. IEEE CVPR, 2004.

[18] J. Chen, D. Yi, J. Yang, G. Zaho, S. Z. Li, M. Pietikainen, Learning Mappings for Face Synthesis from Near Infrared to Visual Light Images, in Proc. IEEE CVPR, 2009, pp. 156-163.

[19] H. Han, S. Shan, X. Chen, S. Lao, W.Gao, Separability Oriented Preprocessing for Illumination-Insensitive Face Recognitio, in Proc. ECCV, 2012, pp. 307-320.

[20] H. Han, S. Shan, X. Chen, W.Gao, A Comparative Study on Illumination Preprocessing in Face Recognition, Pattern Recognit. 46 (2013), pp. 1691-1699.

[21] J.-S. Lee, Digital Image Enhancement and Noise Filtering by Use of Local Statistics, IEEE Trans. Pattern Anal. Mach. Intell. 2 (1980) pp. 165-168.

[22] R. Wang, J. Yang, D. Yi, S. Z. Li, An Analysis-by-Synthesis Method for Heterogeneous Face Biometrics, in Proc. ICB, 2009, pp. 319-326.

[23] Z. Zhang, Y. Wang, Z. Zhang, Face Synthesis from Near-Infrared to Visual Light Via Sparse Representation, in Proc. IJCB, 2011, pp. 1-6.

[24] D. Yi, R. Liu, R. Chu, Z. Lei, S. Z. Li, Face Matching between Near Infrared and Visible Light Images, in Proc. ICB, 2007, pp. 523-530.

[25] B. Klare, A. K. Jain, Heterogeneous Face Recognition: Matching NIR to Visible Light Images, in Proc. ICPR, 2010, pp. 1513-1516.

[26] H. Maeng, S. Liao, D. Kang, S.-W. Lee, A. K. Jain, Nighttime Face Recognition at Long Distance: Cross-distance and Cross-spectral Matching, in Proc. ACCV, 2012, pp. 708-721.

[27] FaceVACS Software Developer Kit, Cognitec Systems GmbH, `http://www.cognitec-systems.de`.

[28] PittPatt Software Developer Kit, Pittsburgh Pattern Recognition, Inc.

[29] B. F. Klare, A. K. Jain, Heterogeneous Face Recognition using Kernel Prototype Similarities, IEEE Trans. Pattern Anal. Mach. Intell. 35 (2013) pp. 1410-1422.

[30] S. Liao, D. Yi, Z. Lei, R. Qin, S. Z. Li, Heterogeneous Face Recognition from Local Structures of Normalized Appearance, in Proc. ICB, 2009, pp. 209-218.

[31] S. T. Roweis, L. K. Saul, Nonlinear Dimensionality Reduction by Locally Linear Embedding, Science 290 (2000) pp. 2323-2326.

[32] A. K. Jain, R. C. Dubes, Algorithms for Clustering Data, Prentice-Hall, 1988.

[33] D. Lowe, Distinctive Image Features from Scale-Invariant Keypoints, Int. J. Comput. Vision 60 (2004) pp. 91-110.

[34] B. Klare, Z. Li, A. K. Jain, Matching Forensic Sketches to Mugshot Photos, IEEE Trans. Pattern Anal. Mach. Intell. 33 (2011) pp. 639-646.

[35] S. Z. Li, Z. Lei, M. Ao, The HFB Face Database for Heterogeneous Face Biometrics Research, in Proc. IEEE CVPR Workshops, pp. 1-8, 2009.

[36] Y. Li, S. Shan, H. Zhang, S. Lao, X. Chen, Fusing Magnitude and Phase Features for Robust Face Recognition, in Proc. ACCV, pp. 601-612, 2012.

[37] H. Han, B. Klare, K. Bonnen, A. K. Jain, Matching Composite Sketches to Face Photos: A Component-Based Approach, IEEE Trans. Inf. Forensics Security, 8 (2013) pp. 191-204.

[38] Near-infrared Illuminator, Raytec, `http://www.rayteccctv.com`.

[39] D. S. Bolme, J. R. Beveridge, M. L. Teixeira, B. A. Draper, The CSU Face Identification Evaluation System: Its Purpose, Features and Structure, in Proc. ICVS, pp. 304-313, 2003.

[40] Q. Liu, X. Tang, H. Jin, H. Lu, S. Ma, A Nonlinear Approach for Face Sketch Synthesis and Recognition, in Proc. IEEE CVPR, 2005, pp. 1005-1010.

[41] K. Ricanek, T. Tesafaye, MORPH: A longitudinal image database of normal adult age-progression, in Proc. FGR, 2006, pp.341-345.

[42] W. Gao, B. Cao, S. Shan, X. Chen, D. Zhou, X. Zhang, D. Zhao, The CAS-PEAL Large-Scale Chinese Face Database and Baseline Evaluations, IEEE Trans. Syst. Man Cybern. Paart A-Syst. Hum. 38 (2013) pp. 149-161.

[43] CASIA-FaceV5, `http://biometrics.idealtest.org/`.

[44] X. Wang, X. Tang, Face Photo-Sketch Synthesis and Recognition, IEEE Trans. Pattern Anal. Mach. Intell. 31 (2009) pp. 1955-1967.

[45] C. Cao, Y. Weng, S. Zhou, Y. Tong, K. Zhou, FaceWarehouse: A 3D Facial Expression Database for Visual Computing, IEEE Trans. Vis. Comput. Graph. 20 (2014) pp. 413-425.