2017 IEEE 12th International Conference on Automatic Face & Gesture Recognition

Heterogeneous Face Recognition: Recent Advances in Infrared-to-Visible Matching

Shuowen Hu[1], Nathaniel Short[2], Benjamin S. Riggan[1], Matthew Chasse[2], M. Saquib Sarfraz[3]
[1]U.S. Army Research Laboratory, Adelphi, MD 20783, USA
[2]Booz Allen Hamilton, McLean, Virginia 22102, USA
[3]Institute of Anthropomatics & Robotics, Karlsruhe Institute of Technology, Karlsruhe, Germany

***Abstract*—An emerging topic in face recognition is matching between facial images acquired from different sensing modalities, referred to as heterogeneous face recognition. Heterogeneous face recognition has the potential to provide key capabilities for the commercial sector as well as for law enforcement, intelligence gathering, and the military, especially in challenging unconstrained settings. However, the difficulty in heterogeneous face recognition is compounded by phenomenology differences between modalities, giving rise to significant facial appearance variations due to the modality gap. In this paper, we focus on a subset of heterogeneous face recognition and present a succinct review of recent work on infrared-to-visible face recognition.**

I. INTRODUCTION (*HEADING 1*)

Face recognition research and development have focused primarily on the visible spectrum in the past few decades, attempting to address challenges in expression, illumination, pose, and resolution. Since the seminal eigenface approach was developed by Sirovich and Kirby in 1987 [1], and later expanded by Turk and Pentland in 1991 [2], significant advancements have been made for face recognition in the visible spectrum, addressing challenges such as illumination, expressions, pose, and resolution. The availability of low-cost digital cameras and the ubiquitous cell phone camera have led to a surge in the amount of visible imagery captured across the world. Coupled with the increasing popularity of social network sites (e.g. Facebook, Twitter) and media sharing websites (e.g. Youtube), there is a massive amount of visible face imagery acquired under different conditions/settings with different visible camera models that can be used to train complex algorithms such as deep neural networks to accurately detect and recognize faces. Infrared cameras, on the other hand, are usually significantly more expensive than their visible counterpart, and acquire imagery in spectral bands not perceivable by the human visual system. Consequently, the use of infrared imagers have been traditionally for military applications in target detection and recognition, especially using thermal infrared sensors. However, there is now an emerging interest in using infrared sensors for biometric face recognition. One of the motivations behind utilizing infrared for face image capture is illumination invariance, which is a significant confound for visible face recognition [4], especially in unconstrained settings

The infrared spectrum consists of four main bands: near infrared (NIR, 0.75-1.4 μm), short-wave infrared (SWIR, 1.4-3 μm), mid-wave infrared (MWIR, 3-5 μm), and long-wave infrared (LWIR, 8-15 μm). In comparison, the visible spectrum is the portion of the electromagnetic spectrum that is perceivable by the human visual system, and includes wavelengths from 0.4 μm to 0.75 μm. The NIR and SWIR bands compose the reflection-dominated region of the infrared spectrum (phenomenology in the visible spectrum is also reflection-dominated), while the MWIR and LWIR bands compose the emission-dominated region (and are collectively referred to as thermal infrared). Figure 1 shows images of a subject in the visible and infrared bands. As can be observed, the NIR and SWIR images more closely resemble the visible spectrum image than the MWIR and LWIR imagery. This is expected, as imaging in the visible, NIR, and SWIR bands acquires reflected radiation, while imaging in the thermal bands acquires radiation mainly emitted from facial skin tissue. Therefore, matching infrared imagery to visible spectrum imagery increases in difficulty as the wavelength increases in the infrared spectrum due to the enlarging modality gap.

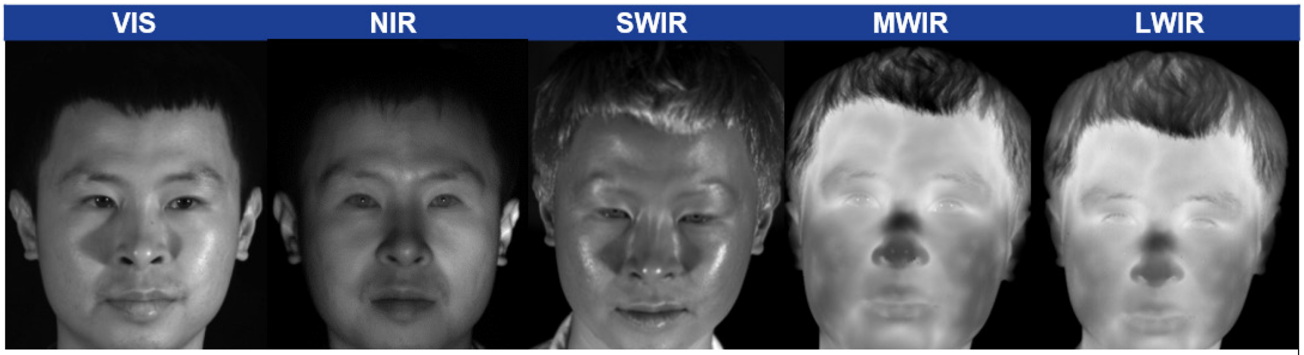

Figure 1. Visible and infrared imagery of a subject.

978-1-5090-4023-0/17 $31.00 © 2017 IEEE
DOI 10.1109/FG.2017.126

Infrared-to-visible matching, the focus of this paper, is a subset of the broader heterogeneous face recognition research area.

The definition of heterogeneous face recognition (HFR) is the matching of probe face imagery acquired in one modality or domain to gallery face imagery acquired in another modality or domain. There are several useful biometric applications that motivates HFR. One of the earliest scenarios for HFR is sketch-to-photo face recognition [3], where a hand-drawn face image is matched to a gallery of visible photo images. The primary motivation is to aid law enforcement in finding a suspect whose appearance was sketched by an artist from a description provided by a witness or victim. Other examples of HFR scenarios are low-resolution to high resolution matching [41], video-to-still matching [42], and 3D-to-2D matching [43]. In this paper, we focus on the infrared-to-visible matching scenario. Note that while the converse of infrared-to-visible face recognition (i.e. visible-to-infrared face recognition; matching a visible probe image to an infrared gallery set) is equally valid as a HFR scenario, we emphasize infrared-to-visible matching because of its operational relevance. Since, most existing government watch lists and biometric databases only contain visible face imagery of individuals of interest, the concept of operations is therefore to match an infrared probe image to a gallery of visible images.

This paper will discuss four HFR scenarios (NIR-to-VIS, SWIR-to-VIS, MWIR-to-VIS, and LWIR-to-VIS), providing an overview of each infrared-to-visible matching scenario and reviewing the first studies published on each scenario as well as representative literature published in recent years. Section II will present a brief correlation experiment showing the increasing challenge of infrared-to-visible face recognition as wavelength increases. Sections III.B and III.C present the NIR-to-VIS and SWIR-to-VIS face recognition scenarios, covering the reflective part of the infrared spectrum. Sections IV.A and IV.B focus on the MWIR-to-VIS and LWIR-to-VIS face recognition scenarios, covering the emissive part of the infrared spectrum. Section IV.C describes the use of polarimetric information in the LWIR band to enhance HFR performance over conventional LWIR-to-VIS face recognition. Section V concludes the paper.

II. Modality Gap

To illustrate the degree of difficulty of infrared-to-visible face recognition, especially as the wavelength increases in the infrared, an analysis using structural similarity is conducted between each infrared band and the visible spectrum. Structural similarity between images x and y is defined as

$$SSIM(x,y) = \frac{(2\mu_x\mu_y+C_1)(2\sigma_{xy}+C_2)}{(\mu_x^2+\mu_y^2+C_1)(\sigma_x^2+\sigma_y^2+C_2)}, \qquad (1)$$

where μ_x and μ_y denote the means of the respective images (or local regions), and σ_x and σ_y denote the standard deviations of the respective images (or local regions), and σ_{xy} denotes the cross-covariance.

The images in Figure 1 are first aligned using three fiducial points (centers of left eye and right eye, and base of the nose), and cropped. Note that the pixel intensities in each image have been normalized to [0 255]. Next, the structural similarity (SSIM) formulation of [17] is used as a quantitative measure of the modality gap between each infrared face image and the "reference" visible face image. The SSIM formulation of [17] defines the structural information as the attributes that represent the structures of objects in an image, independent of the average luminance and contrast. We compute the SSIM of each infrared image in Figure 1 to the visible reference image of Figure 1, yielding SSIM values of 0.581 for NIR-to-VIS, 0.491 for SWIR-to-VIS, 0.368 for MWIR-to-VIS, and 0.335 for LWIR-to-VIS. As the wavelength in the infrared band increases, the modality gap (represented here by the SSIM) also increases, with a sharp decrease in SSIM transitioning from the reflective infrared region to the emissive infrared region. Figure 2 shows the SSIM maps (computed for each pixel using a circular-symmetric window with standard deviation of 1.5 samples [17]). Note that the infrared images used here were acquired in the same trial/session as the visible reference image. As a comparison, we also computed the SSIM using a visible image of the same subject from another trial with the reference visible image, yielding a SSIM value of 0.751, even though the pose of the subject have changed a few degrees between sessions.

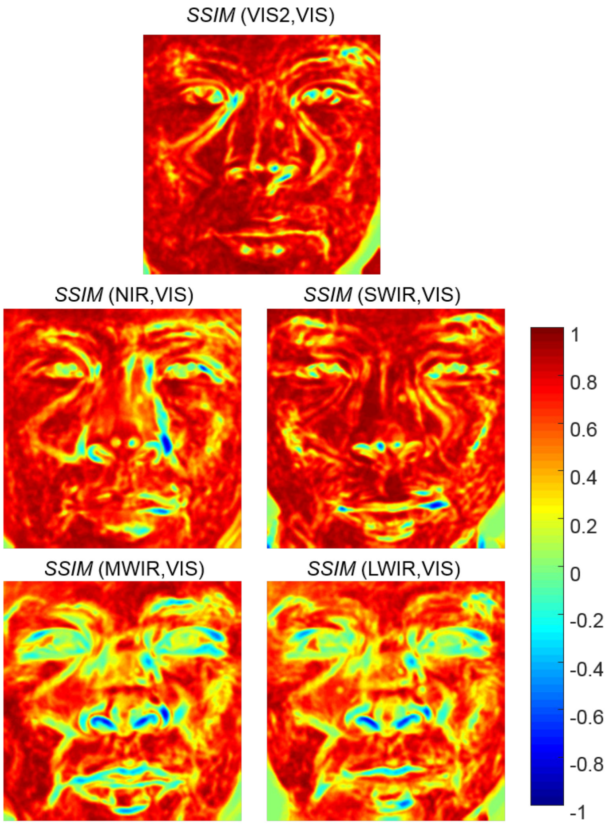

Figure 2. Structural similarity maps between infrared image and reference visible image.

As can be observed in Figure 1, in the higher frequency regions of the face (around the eyes, nose, and mouth), there are increasing differences between the infrared and visible face signatures as wavelength increases in the infrared. Even in the smoother facial regions (e.g. cheeks), the temperature variations in skin tissue cause those regions to be more dissimilar between the thermal bands and the visible reference, as expected. Therefore, the degree of HFR difficulty increases from the NIR-to-VIS face recognition scenario to the hardest LWIR-to-VIS scenario.

III. Reflective IR-to-Visible Face Recognition

A. *NIR-to-Visible*

NIR-to-visible face recognition is one of the earlier HFR scenarios to be conceived, after sketch-to-photo. The benefit of NIR imaging is its relative invariance to environmental lighting compared to visible imaging, though an active NIR illuminator are often employed, especially for low-light or nighttime imaging. For short distance imaging (the most common application), NIR illuminators in the form of LEDS are typically used, which are safe to the naked eye. However, for long distance imaging, NIR illuminators in the form of lasers with wavelengths in the 0.81 µm to 0.94 µm range are typically employed, which may cause photothermal damage in the retina [5], depending on the intensity level of the laser beam when reaching the eye. The NIR wavelengths are generally not perceivable by the human visual system, offering some degree of covertness for nighttime operation. However, NIR illuminators can be readily observed by silicon-based image sensors and, in fewer cases, the human eye [6].

The earliest work on NIR-to-VIS face recognition is by Yi et al. [7], who formulated the problem as correlational regression between an NIR face image and a visible face image which have been aligned by the eye coordinates and normalized. [7] proposed a three step process for NIR-to-VIS matching: extracting lower-dimensionality features from the NIR and VIS imagery, then performing multivariate regression between the features, and finally evaluating the similarity score based on a correlation score. For the first step, principal component analysis (PCA) and linear discriminant analysis (LDA) were used for feature extraction and dimensionality reduction. Next, canonical correlation analysis (CCA) was used to compute the best correlational regression between the features vectors of the images extracted from the two different modalities. The final step computes the correlation score after projection of the NIR and VIS feature vectors into CCA subspace. On a dataset of 200 subjects, [7] achieved a verification rate of 93.1% at false alarm rate (FAR) of 0.1% [7].

To support algorithm development for NIR based face recognition, several databases have been collected by universities and are publicly available through request. Table I lists four of the most extensive databases available to researchers. [11] contains only NIR face imagery, while [8-10] contain both NIR and visible spectrum face images of each subject, and are suitable for HFR research.

Since the initial work of [7], a number of research groups have further advanced NIR-to-VIS face recognition. Here, we discuss several recent works that are representative state-of-the-art approaches. The advances in visible face recognition through deep learning have been made possible due to two main factors: advances in computing technology, and the massive amounts of visible spectrum face imagery that are available on the Internet. Though there are several NIR databases, the amount of imagery contained in these databases do not rival the millions of visible face images that are frequently used to train deep neural networks for visible face recognition. Given that the modality gap between visible and NIR imaging is relatively small, Liu et al. [12] leverages recent deep learning advances by pre-training a deep CNN on the visible imagery from the large-scale CASIA WebFace Database [13], and then fine-tuning the model on NIR face imagery to learn a domain-invariant deep representation. [12] used a triplet formulation with two types of NIR-VIS triplet loss to reduce intra-class variations and augment the number of positive sample training pairs. By performing hard sample selection, their technique achieved a verification rate of 91.03% at FAR of 0.1% and Rank-1 identification accuracy of 95.74% on the CASIA NIR-VIS 2.0 database. In the same spirit, Reale et al. [14] utilized deep networks with small convolutional filters, and pre-trained on the visible CASIA WebFace Database. For NIR-to-VIS face recognition, [14] initialized two networks based on the pre-trained network, excluded the softmax classifier and removed the fully connected layer, and trained these two networks (named VisNet and NIRNet), coupling their output features by creating a Siamese network with contrastive loss. [14] achieved a Rank-1 accuracy of 87.1% and a verification rate of 74.5% at FAR of 0.1%. A more recent similar CNN pipeline has been introduced in [44], where different hyperparameter design choices have led to substantial performance gains on CASIA NIR-Vis 2.0 database achieving 95.82% Rank-1 accuracy and a verification rate of 94.03% at 0.1 FAR.

A different approach to solving NIR-to-visible face recognition is to reconstruct/estimate the visible image corresponding to a NIR input image (i.e. synthesizing VIS from NIR). Conceptually, this approach can serve as a preprocessing step, after which the reconstructed VIS image corresponding to the original NIR probe image can be entered into an existing visible face recognition system. With this approach, government agencies would not need to invest in a separate NIR-to-visible face recognition system, but instead would add the reconstruction software as preprocessing stage to already deployed visible face recognition systems. Juefei-Xu et al. [15] recently proposed cross-spectral joint dictionary learning and reconstruction based on K-SVD. Their algorithm jointly learns a NIR and a VIS dictionary, enforcing that the sparse representation be the same for of the NIR and VIS images in each dictionary [15]. On the CASIA NIR-VIS v2.0 database, [15] achieved Rank-1 accuracy of 78.5% and a verification rate of 85.8% at FAR of 0.1%.

TABLE I. NIR FACE DATABASES

Database Name	**# of Subjects**
CASIA NIR-VIS 2.0 [8]	725
ND-Near Infrared and Visible Light (ND-NIVL) [9]	574
Long Distance Heterogeneous Face Database (LDHF-DB) [10]	100
Hong Kong Polytechnic University (PolyU) NIR Face Database [11]	335

B. *SWIR-to-Visible*

The phenomenology in the SWIR band, like the NIR band, is also reflection dominated. However, the difference in facial signatures between SWIR and visible is more pronounced, creating a more challenging HFR scenario than NIR-to-VIS face recognition. Though imaging SWIR typically requires an active illuminator in low-light or during nighttime, it is not perceivable to the human eye and can be made more covert by illuminating at selected wavelengths [6]. Furthermore, SWIR can more effectively penetrate fog, haze, and dust (i.e. atmospheric obscurants) [16]. A major challenge with SWIR is that eye-safe illumination is above 1.4 µm – however, moisture in the skin absorbs infrared wavelengths above 1.45 µm, causing the facial skin to appear dark in low-light imagery acquired with an eye-safe SWIR illuminator [6].

The amount of research dedicated to face recognition in the SWIR band is very limited, with no available public dataset containing SWIR faces. To the best of our knowledge, [18] was the first work on SWIR-to-VIS face recognition, assessing photometric normalization through contrast limited adaptive histogram equalization (CLAHE) followed by PCA with k-nearest neighbor matching. [18] also assessed several commercial matchers for SWIR-to-VIS face recognition. [19] extended the results of [18] by assessing more photometric normalization methods based on single scale retinex and proposed cross-photometric score level fusion to improve performance. The most recent work by Cao et al. [20] on SWIR-to-visible face recognition proposed composite multilobe descriptors, which combines a Gaussian function with local binary patterns, Weber local descriptor, and histogram of oriented gradients (HOG) into multilobe operators. [20] achieved a verification rate of 99.5% at FAR of 10% and Rank-1 accuracy of 78.7% with SWIR imagery collected at 1.5 m on a private dataset of 48 subjects.

There has not been as many studies performed on the SWIR-to-VIS scenario than the other infrared-to-visible HFR scenarios. This may be partly due to the lack of a publicly available database to support algorithm development by researchers who may not have access to SWIR imagers for in-house data collection. Given that the SWIR facial signature is still somewhat similar to the visible facial signature, we expect that SWIR-to-visible face recognition can exploit deep learning based face recognition techniques in the visible spectrum through fine tuning/transfer learning. However, this is a conjecture based on the reflection dominance in SWIR.

IV. Emissive IR-to-Visible Face Recognition

A. *MWIR-to-Visible*

Imaging in the MWIR band, unlike in the SWIR and NIR bands, is emission dominated, though there is a stronger reflective component in the MWIR than LWIR. Imaging in the MWIR band is completely passive, not requiring active illuminating during daytime or nighttime, and is therefore more covert than imaging in the SWIR and NIR bands that require an active illuminator in low-light and nighttime settings. Human facial skin tissue has high emissivity in the MWIR band (0.91 as measured by Wolff et al. [21]) – the face signature acquired in MWIR reflects the heat distribution arising from the underlying vasculature and depends on an individual's physiology. Consequently, there is a large modality gap between the MWIR and visible human face signatures, rendering MWIR-to-VIS face recognition a highly challenging HFR scenario. The availability of MWIR face databases is also limited – a few databases can be requested, though they are not publicly available. The Pinellas County Sheriff's Office (PCSO) collected MWIR and visible images of 1000 subjects, and was first used in the study of Klare and Jain [22]. The US Army CERDEC-NVESD database collected jointly with the US Army Research Laboratory contains MWIR, LWIR, and visible imagery of 50 subjects, and was first used in the study of Hu et al. [23].

One of the first published work on MWIR-to-VIS face recognition was by Bourlai et al. [24], who evaluated preprocessing, feature extraction, and similarity metrics as a complete processing chain for matching. [24] reported that the best Rank-1 accuracy of 53.9% on an in-house dataset of 39 subjects was achieved with difference-of-Gaussian preprocessing, three patch local binary patterns (LBP) feature extraction, and chi-squared distance based matching. Klare and Jain [22] developed a nonlinear kernel prototype based approach that represents features extracted from heterogeneous image modalities, followed by linear discriminant analysis (LDA) to improve the discriminative capabilities of the prototype representations. Testing on a subset (333 subjects) of the PCSO dataset and augmenting the visible gallery with 10,000 additional subjects, [22] achieved a verification rate of 78.2% at FAR=1%. Hu et al. [23] developed an approach using DOG filtering, followed by HOG feature extraction, and partial least squares (PLS) based matching. The incorporation of thermal cross-examples as negative samples in the PLS framework improved recognition performance, achieving a verification rate of 94.8% at FAR=1% on a 48-subject gallery using the NVESD dataset.

More recently, Chen and Ross [25] proposed a cascaded subspace learning framework consisting of whitening transformation, factor analysis, and common discriminant analysis, seeking to extract identity features that are invariant across spectral bands. First, [25] reduced cross-spectral differences in facial signatures through photometric adjustment. Then, histograms of principal oriented gradients and a variant of the scale invariant feature transform (SIFT) called PSIFT are extracted as feature vectors. Next, multiple subspaces are constructed by random sampling of image patches, and the corresponding feature vectors are used as input into the cascaded subspace learning framework. On the PCSO dataset, [25] achieved a verification rate of 80.9% at FAR=1%. Sarfraz and Stiefelhagen [26] proposed a neural network based approach called deep perceptual mapping (DPM) to bridge the modality gap and facilitate MWIR-to-VIS face recognition. [26] used a 3-layer (2 hidden layers) neural network to learn a non-linear mapping between the SIFT features extracted from the DOG filtered facial images in the MWIR and visible domains. This approach can be considered a direct regression approach that maps SIFT features from the MWIR domain to its visible representation (or vice versa), and can be effectively trained on a relatively small training dataset. [26] achieved a high Rank-1 identification accuracy of 98.7% on the NVESD dataset using all available visible imagery per subject to form a 25-subject gallery. Whereas [26] performs direct regression, Riggan et al. [27] proposed a novel coupled autoassociative neural network that performs an indirect regression between visible SIFT features and MWIR SIFT features to extract common latent features. [27] achieved a Rank-1 accuracy of 94.4% on the NVESD dataset with a 40-subject gallery.

B. LWIR-to-Visible

Facial signatures acquired in the MWIR and LWIR bands are visually and phenomenologically similar to a large extent. The emissivity of facial skin tissue is 0.97 in LWIR [21], which is slightly higher than in MWIR. Therefore, techniques developed for MWIR-to-VIS face recognition can also be readily applied to LWIR-to-VIS face recognition. However, LWIR-to-VIS face recognition is a more challenging HFR scenario than MWIR-to-VIS face recognition, due to several factors. Firstly, the resolving power of any imager is limited to the wavelength of the radiation. Since the wavelength is longer in LWIR than MWIR, imagery acquired in LWIR has inherently less spatial resolution. However, since facial structures are not typically on the order of micrometers (the wavelength in the thermal spectrum), this is not a significant limiting factor for LWIR-to-VIS face recognition versus MWIR-to-VIS. There are, however, major differences between the sensors typically used to acquire MWIR and LWIR imagery. MWIR imagers are predominantly cooled systems that integrate the imaging sensor with a cryocooler, which lowers the sensor temperature to cryogenic conditions. LWIR imagers may also be cooled systems using cryocoolers, similar in design to the MWIR imagers. However, a large segment of the commercial market for LWIR imagers relies on uncooled designs, the most common form being the microbolometer [28]. Microbolometers are significantly less expensive than their cooled counterparts, often by an order of magnitude or more. Though more cost effective, microbolometers have lower sensitivity, lower signal-to-noise ratio, and lower spatial resolution (partly due to the typically larger detector pitch). Therefore, LWIR-to-VIS face recognition is the most challenging of the infrared-to-visible heterogeneous face recognition scenarios.

Due to the accessibility and availability of lower cost microbolometers, there are more LWIR face databases available to facilitate HFR research. One of the first LWIR face databases is the 82-subject ND-Collection X1, collected by the University of Notre Dame in the early 2000's [29], which also contained visible imagery. Since then, more LWIR face databases have collected with improved imagers as the sensor technology has rapidly evolved – Table II lists several commonly used databases for LWIR-to-VIS face recognition. Note that thermal imagery (both MWIR and LWIR) can also facilitate the detection of disguises, which is the focus of the database introduced in [32].

The first study published on the LWIR-to-VIS scenario is [33], which evaluated several different algorithms for preprocessing and feature extraction for LWIR-to-VIS face recognition using PLS. On the ND-Collection X1 database, [33] achieved the best Rank-1 identification rate of 49.9% on a testing set of 41 subjects with DOG filtering and HOG feature extraction. Results on the more recent thermal-to-visible face recognition techniques (please refer back to Section IV.A for more algorithm details), are as follows. This method of [33] was extended in [23], improving the Rank-1 accuracy to 72.7% on the 41-subject testing set from the ND-Collection X1 database. On the LWIR portion of the NVESD dataset, [23] achieved a verification rate of 80.2% at FAR=1% on a 48-subject gallery. Note that this verification rate is notably lower than the 94.8% achieved for MWIR-to-VIS matching, demonstrating that LWIR-to-VIS face recognition is more challenging. On the NVESD dataset, [27] achieved Rank-1 accuracy of 89.1% for LWIR-to-VIS face recognition, also lower than the reported 94.4% Rank-1 accuracy for MWIR-to-VIS face recognition. Also on the NVESD dataset, [26] achieved Rank-1 accuracy of 97.3% for LWIR-to-VIS matching, only slightly lower than the 98.7% reported for MWIR-to-VIS matching. On the Carl database, which collected the thermal imagery using a microbolometer, [25] achieved a Rank-1 accuracy of 75.6% and a verification rate of 51.2% at FAR = 1%.

Almost all the databases containing MWIR and LWIR facial imagery are collected in controlled settings, typically at close ranges. The reported results for thermal-to-visible face recognition are therefore for ideal conditions and mostly for frontal face imagery, matching against limited gallery sizes. Even so, the HFR performance (in terms of both identification and verification accuracy) has only achieved limited success, illustrating the significant challenge for thermal-to-visible face recognition. Overcoming these challenges will lead to a critical capability that can provide covert day and night face recognition.

TABLE II. LWIR FACE DATABASES

Database Name	# of Subjects
ND-Collection X1 [29]	82
Carl Database [30]	41
OTCBVS Dataset 02 [31]	30
IIIT-Delhi In and Beyond Visible Spectrum Disguise Database [32]	75
NVESD Database [23]	50

C. Polarimetric LWIR-to-Visible

An emerging area of research is the use of polarization state information of LWIR emissions to facilitate face recognition. Note that polarization state information can be measured in any infrared band as well as in the visible spectrum – here, we focus on the LWIR band. The polarization states are described using the Stokes parameters $\boldsymbol{S}_0$, $\boldsymbol{S}_1$, and $\boldsymbol{S}_2$ [34], which are used to compute the degree of linear polarization (DoLP) traditionally used to visualize polarimetric imagery. The Stokes parameters are derived by measuring radiant intensities of the linear states of polarization at angles of 0°, 45°, 90°, and 135°. Gurton et al. [35] is the first study that presented polarimetric LWIR facial imagery. Figure 3 shows polarimetric imagery of a subject. $\boldsymbol{S}_0$ represents the conventional LWIR image without any polarization, while $\boldsymbol{S}_1$ represents horizontal & vertical polarization, and S_2 represents the diagonal polarization. As can be observed, the polarization state information in S_1 and S_2 contain geometric and textural facial details that are not present in S_0, and can be used to complement S_0 in improving face recognition performance. Furthermore, the Stokes images can be used to estimate the surface normal at each pixel, enabling a 3D facial surface to be reconstructed [36] which can potentially be used to "frontalize" off-angle probe images to match against galleries containing predominantly frontal face imagery.

[37] is the first study to combine edge orientation features extracted from S_0, S_1, and S_2, demonstrating that polarimetric LWIR-to-VIS face recognition outperformed conventional LWIR-to-VIS face recognition. The most recent work by Riggan et al. [38] proposed an optimal feature learning and discriminative framework for polarimetric LWIR-to-VIS face recognition, extending the coupled autoassociative neural network followed by PLS to improve the recognition accuracy. [38] reported Rank-1 identification rate of 93.3% on a 50-subject database split into equal subsets for training and testing. An extended version of that database containing 60 subjects is now available upon request (please refer to [39]). Figure 4 shows the cumulative match characteristic curves on this extended database comparing polarimetric LWIR-to-VIS to conventional LWIR-to-VIS face recognition (treating S_0 as the conventional LWIR probe set), using couple neural networks followed by PLS and a visible face matcher, PittPatt SDK 5.2.2. Figure 4 illustrates that polarization state information helps improve HFR performance using the approach of [38], which extracted SIFT features, followed by PCA for dimensionality reduction, indirect regression through CpNN, and PLS for classification. Figure 4 also illustrates that PittPatt cannot overcome the large modality gap for polarimetric and conventional LWIR-to-VIS.

More recently, Riggan et al. [40] introduced a method of estimating/synthesizing a visible spectrum face image from a polarimetric LWIR face image. [40] used a two-step approach, first mapping the SIFT features extracted from a polarimetric thermal image to its visible feature representation, from which the corresponding visible image can be reconstructed using a convolutional neural network based approach. The advantage of such a synthesis approach is its ability to provide a preprocessing stage to existing visible face matchers for heterogeneous face recognition.

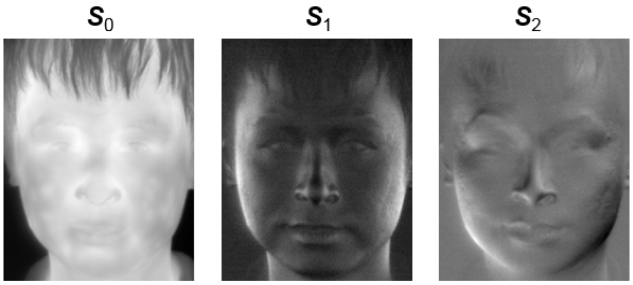

Figure 3. Polarimetric LWIR Stokes images of a subject.

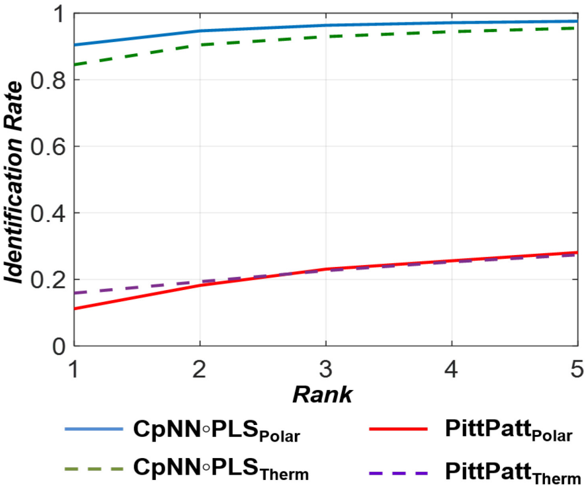

Figure 4. Cumulative match characteristic curves for polarimetric LWIR-to-VIS and conventional LWIR-to-VIS.

V. Discussion and Conclusion

In this paper, we discussed and reviewed recent works on heterogeneous face recognition, focusing on the infrared-to-visible matching scenarios. The reflective infrared region (NIR, SWIR) offers some advantages for face recognition such as imaging through fog and haze, and in low-light conditions with an active illuminator not observable to the human eye. Face signatures in these bands are also more similar to the visible face signature, allowing researchers to potentially adapt state-of-the-art deep learning based approaches trained on visible imagery for heterogeneous face recognition. Imaging in the emissive infrared region (MWIR and LWIR) has the advantage of being purely passive, offering a truly covert surveillance capability. However, due to the difference in facial signatures between the thermal spectrum and the visible spectrum, heterogeneous face recognition is much more challenging. Polarimetric imaging in the thermal spectrum provides additional geometric and textural details that facilitate matching with visible spectrum imagery. Though significant progress has been made in infrared-to-visible face recognition, researchers in HFR are impeded by a lack of a large multi-modal face database. The collection of such a database on the order of a thousand subjects, containing imagery simultaneously acquired across the infrared bands under various conditions and settings, will greatly facilitate HFR research. Furthermore, such a large multi-modal face database will also facilitate algorithm development for automated face detection and fiducial point (e.g. eyes, nose, mouth, etc.) labeling algorithms, which are important processing stages prior to the actual matching/recognition stage.

With regards to algorithm development for infrared-to-visible face recognition, transfer learning based approaches leveraging existing deep neural networks trained on large amounts of visible face imagery are effective for the NIR-to-VIS scenario, and likely to be effective for the SWIR-to-visible scenario as well (though, to the best of our knowledge, no transfer learning studies have been conducted on SWIR-to-visible HFR due to the lack of a publicly available SWIR face database). Facial signatures in the NIR and SWIR bands are more similar to the corresponding visible face signature, as the phenomenology is reflection dominated in these bands. However, due to the much wider modality gap arising from phenomenology differences, we conjecture that transfer learning will not be as effective unless the thermal images are first brought closer to the visible counterpart via some prior learned functional mappings at the pixel level. For thermal-to-visible HFR, regression/mapping techniques via relatively shallow neural networks relying on handcrafted edge-based features like SIFT or HOG are the most effective at the present, given the limited availability of face data in the MWIR and LWIR bands. We believe that approaches using convolutional neural networks that do not rely on handcrafted features as input is the path forward for thermal-to-visible HFR. However, this would likely entail additional data to be collected and made available to the community.

Infrared-to-visible HFR, as well as HFR in general, is still a nascent research area. To reiterate, the collection of a large scale face data across all the spectral bands under varying conditions (e.g. range, pose, etc.) would greatly benefit the development of algorithms from face detection to fiducial point labeling to recognition. As the HFR area continues to develop, we expect that new HFR systems will enable key capabilities for commercial, military, and law enforcement applications, providing interoperability with visible spectrum face imagery in existing biometric watch lists and social media sites.

References

[1] L. Sirovich and M. Kirby, "Low-dimensional procedure for the characterization of human face," Journal of the Optical Society of America A, vol. 4(3), pp. 519-524, 1987.

[2] M. Turk and A. Pentland, "Face recognition using eigenfaces," *Proc. IEEE Int'l Conf. Computer Vision and Pattern Recognition,* pp. 586-591 *1991.*

[3] R.G. Uhl and N.D.V. Lobo, "A framework for recognizing a facial image from a police sketch," *Proc. IEEE Int'l Conf. Computer Vision and Pattern Recognition, 1996.*

[4] Y. Adini, Y. Moses, and S. Ullman, "Face recognition: the problem of compensating for changes in illumination direction," IEEE Trans. Pattern Analysis and Machine Intelligence, vol. 19(7), pp. 721-732, 1997.

[5] W. Calhoun, "Near infrared (NIR) illuminators for surveillance," presented at the *Technical Electronic Product Radiation Safety Standards Committee Meeting,* Oct. 25, 2016.

[6] J.M Dawson, S.C. Leffel, C. Whitelam, T. Bourlai, "Collection of multispectral biometric data for cross-spectral identification applications," in Face Recognition Across the Imaging Spectrum, Ed. T. Bourlai, pp. 21-46, 2016.

[7] D. Yi, R. Liu, R. Chu, Z. Lei, S.Z.Li, "Face matching between near infrared and visible light images," *Proc. Int'l Conf. on Biometrics,* pp. 523-530, 2007.

[8] S.Z. Li, D. Yi, Z. Lei, S. Liao, "The CASIA NIR-VIS 2.0 Face Database," *Proc. IEEE Workshop on Perception Beyond the Visible Spectrum,* 2013.

[9] J. Bernhard, J. Barr, K.W. Bowyer, P.J. Flynn, "Near-IR to visible light face matching: Effectiveness of pre-processing options for commercial matchers," *Proc. IEEE Int'l Conf. on Biometrics: Theory, Applications, and Systems,* 2015.

[10] D. Kang, H. Han, A.K. Jain, S.-W. Lee, "Nighttime face recognition at large standoff: cross-distance and cross-spectral matching," Pattern Recognition, vol. 47(12), pp.3750-3766, 2014.

[11] B. Zhang, L. Zhang, D. Zhang, L. Shen, "Directional binary code with application to PolyU near-infrared face database," Pattern Recognition Letters, vol. 31(14), pp. 2337-2344, 2010.

[12] X. Liu, L. Song, X. Wu, T. Tan, "Transferring deep representation for NIR-VIS heterogeneous face recognition," *Proc. Int'l Conf. on Biometrics,* 2016.

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

[14] C. Reale, N.M. Nasrabadi, H. Kwon, R. Chellappa, "Seeing the forest from the trees: a holistic approach to near-infrared heterogeneous face recognition," *Proc. IEEE Workshop on Perception Beyond the Visible Spectrum,* 2016.

[15] F. Juefei-Xu, D.K. Pal, M. Savvides, "NIR-VIS heterogeneous face recognition via cross-spectral joint dictionary learning and reconstruction," *Proc. Int'l Conf. on Computer Vision and Pattern Recognition Workshops,* 2015.

[16] Goodrich (UTAS): Defense File, 06 April 2010 [Online]. http://www.defensefile.com/News_Detail_Lightweight_swir_sensor_for_target_detection_on_board_uav_equipment_7430.asp

[17] Z. Wang, A.C. Bovik, H.R. Sheikh, E.P. Simoncelli, "Image quality assessment: from error visibility to structural similarity," IEEE Trans. Image Processing, vol. 13(4), 2004.

[18] T. Bourlai, N.D. Kalka, A. Ross, B. Cukic, L. Hornak, "Cross-spectral face verification in the short wave infrared (SWIR) band," *Proc. Int'l Conf. on Pattern Recognition*, 2010.

[19] N.D. Kalka, T. Bourlai, B. Cukic, L. Hornak, "Cross-spectral face recognition in heterogeneous environments: A case study on matching visible to short-wave infrared imagery," *Proc. Int'l Joint Conf. on Biometrics*, 2011.

[20] Z. Cao, N.A. Schmid, T. Bourlai, "Composite multilobe descriptors for cross-spectral recognition of full and partial face," Optical Engineering, vol. 55(8), 083107, 2016.

[21] L.B. Wolff, D.A. Socolinsky, C.K. Eveland, "Face recognition in the thermal infrared," in Computer Vision Beyond the Visible Spectrum, pp. 167 -191, Springer, 2007.

[22] B. F. Klare and A. K. Jain, "Heterogeneous face recognition using kernel prototype similarity," IEEE Trans. Pattern Anal. Mach. Intell., vol. 35, pp. 1410–1422, 2013.

[23] S. Hu, J. Choi, A.L. Chan, W.R. Schwartz, "Thermal-to-visible face recognition using partial least squares," Journal of the Optical Society of America A, vol. 32(3), pp. 431-442, 2015.

[24] T. Bourlai, A. Ross, C. Chen, and L. Hornak, "A study on using mid-wave infrared images for face recognition," *Proc. SPIE 8371*, 83711K, 2012.

[25] C. Chen, A. Ross, "Matching thermal to visible face images using hidden factor analysis in a cascaded subspace learning framework," Pattern Recognition Letters, vol. 72, pp. 25-32, 2016.

[26] M.S. Sarfraz, R. Stiefelhagen, "Deep perceptual mapping for cross-modal face recognition," International Journal of Computer Vision, 2016.

[27] B.S. Riggan, C. Reale, N.M. Nasrabadi, "Coupled auto-associative neurla networks for heterogeneous face recognition," IEEE Open Access, vol. 3, pp. 1620-1632, 2015.

[28] D. Ostrower, "Optical thermal imaging – replacing microbolometer technology and achieving universal deployment," III-Vs Review, vol. 19(6), pp. 24-27, 2006.

[29] X. Chen, P.J. Flynn, K.W. Bowyer, "Visible-light and infrared face recognition," *Proc. ACM Workshop on Multimodal User Authentication*, pp. 48-55, 2003.

[30] V. Espinosa-Duró, M. Faundez-Zanuy, J. Mekyska, "A new face database simultaneously acquired in visible, near-infrared and thermal spectrums", Cognitive Computation, vol. 5, no. 1, pp. 119-135, 2013.

[31] OTCBVS Benchmark Dataset Collection, http://vcipl-okstate.org/pbvs/bench/. Accessed 08 Feb 2017.

[32] T. I. Dhamecha, A. Nigam, R. Singh, and M. Vatsa, "Disguise Detection and Face Recognition in Visible and Thermal Spectrums," *Proc. International Conference on Biometrics*, 2013.

[33] J. Choi, S. Hu, S. S. Young, and L. S. Davis, "Thermal to visible face recognition," Proc. SPIE 8371, 83711L, 2012.

[34] M. Born and E. Wolf, Principles of Optics (Pergamon, 1959).

[35] K. P. Gurton, A. J. Yuffa, and G. W. Videen, “Enhanced facial recognition for thermal imagery using polarimetric imaging,” Opt. Lett. 39, 3857, 2014.

[36] A. J. Yuffa, K. P. Gurton, and G. W. Videen, “Three-dimensional facial recognition using passive long-wavelength infrared polarimetric imaging,” Applied Optics, vol. 53(36), pp. 8514-8521, 2014.

[37] N. Short, S. Hu, P. Gurram, K. Gurton, A. Chan, “Improving cross-modal face recognition using polarimetric imaging,” Optics Letters, vol. 40(6), pp. 882-885, 2015.

[38] B.S. Riggan, N.J. Short, S. Hu, “Optimal feature learning and discriminative framework for polarimetric thermal to visible face recognition,” Proc. IEEE Winter Conference on Applications of Computer Vision, 2016.

[39] S. Hu, N.J. Short, B.S. Riggan, C. Gordon, K.P. Gurton, M. Thielke, P. Gurram, A.L. Chan, “A polarimetric thermal database for face recognition research,” Proc. IEEE Conference on Computer Vision and Pattern Recognition Workshops, 2016.

[40] B.S. Riggan, N.J. Short, S. Hu, H. Kwon, “Estimation of visible spectrum faces from polarimetric thermal faces,” Proc. IEEE Int’l Conf. on Biometrics Theory, Application and Systems, 2016.

[41] W.W.W. Zou, P.C. Yuen, “Very low resolution face recognition problem,” IEEE Trans. on Image Processing, vol. 21(1), pp. 327-340, 2012.

[42] S. Chen, S. Mau, M.T. Harandi, C. Sanderson, A. Bigdeli, B.C. Lovell, “Face recognition from still images to video sequences: a local feature-based framework,” EURASIP Journal on Image and Video Processing, 2010.

[43] I.A. Kakadiaris, G. Toderic, G. Evangelopoulos, G. Passalis, D. Chu, X. Zhao, S.K. Shah, T. Theharis, “3D-2D face recognition with pose and illumination normalization,” Computer Vision and Image Understanding, vol. 154, pp. 137-151, 2017.

[44] R. He, X. Wu,Z. Sun, T.Tan, “Learning Invariant Deep Representation for NIR--VIS Face Recognition”, AAAI 2017.