

Bridging the spectral gap using image synthesis: a study on matching visible to passive infrared face images

Nnamdi Osia & Thirimachos Bourlai

Machine Vision and Applications

ISSN 0932-8092

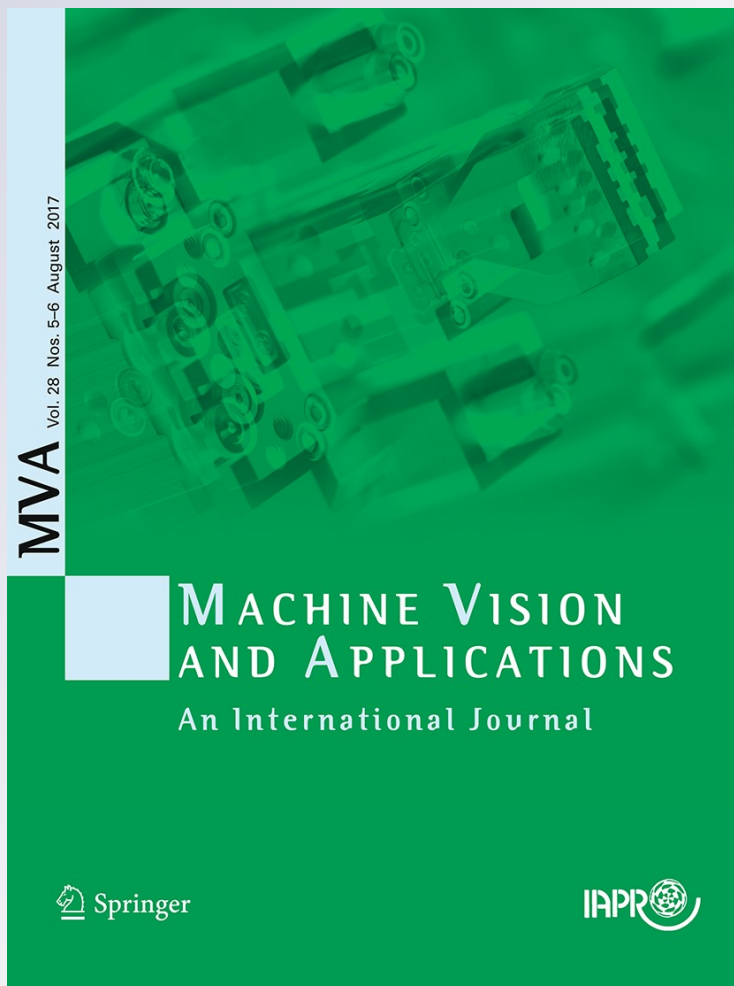
Volume 28

Combined 5-6

Machine Vision and Applications (2017)

28:649-663

DOI 10.1007/s00138-017-0855-1



Your article is protected by copyright and all rights are held exclusively by Springer-Verlag GmbH Germany. This e-offprint is for personal use only and shall not be self-archived in electronic repositories. If you wish to self-archive your article, please use the accepted manuscript version for posting on your own website. You may further deposit the accepted manuscript version in any repository, provided it is only made publicly available 12 months after official publication or later and provided acknowledgement is given to the original source of publication and a link is inserted to the published article on Springer's website. The link must be accompanied by the following text: "The final publication is available at link.springer.com".

Bridging the spectral gap using image synthesis: a study on matching visible to passive infrared face images

Nnamdi Osia¹ · Thirimachos Bourlai¹

Received: 26 November 2016 / Revised: 1 June 2017 / Accepted: 6 June 2017 / Published online: 23 June 2017
© Springer-Verlag GmbH Germany 2017

Abstract We propose an approach that bridges the gap between the visible and IR band of the electromagnetic spectrum, namely the mid-wave infrared or MWIR (3–5 μm) and the long-wave infrared or LWIR (8–14 μm) bands. Specifically, we investigate the benefits and limitations of using synthesized visible face images from thermal and vice versa, in cross-spectral face recognition systems when utilizing canonical correlation analysis and manifold learning dimensionality reduction. There are four primary contributions of this work. First, we assemble a database of frontal face images composed of paired VIS-MWIR and VIS-LWIR face images (using different methods for pre-processing and registration). Second, we formulate a image synthesis framework and post-synthesis restoration methodology, to improve face recognition accuracy. Third, we explore cohort-specific matching (per gender) instead of blind-based matching (when all images in the gallery are matched against all in the probe set). Finally, by conducting an extensive experimental study, we establish that the proposed scheme increases system performance in terms of rank-1 identification rate. Experimental results suggest that matching visible images against images acquired with passive infrared spectrum, and vice-versa, are feasible with promising results.

Keywords Face recognition · Heterogeneous · Cross-spectral · Visible · Long-wave · Middle-wave · Infrared · Synthesis · Restoration

1 Introduction

Over the last few decades, there has been a concerted effort on exploring face recognition (FR) research for a number of military and law enforcement applications. However, a vast majority of the research related to FR is based on images captured within the visible band (380–750 nm). The biggest challenge with FR is the acquisition of face images under conditions which are not controlled, introducing variation in pose, expression, and illumination. In environments where visibility may be unpredictable and uncontrollable such as during the night time, the sole use of images from a single spectrum (e.g., visible) may not be a viable approach [4, 29, 40]. It is important that there is a push to study and address this heterogeneous FR challenge, using the infrared (IR) spectrum for its benefits [32, 42, 43].

Differences in appearance arise between images sensed in the visible and the active IR band, primarily due to the properties of the object being imaged. The active IR spectrum is composed of the near IR band (0.7–0.9 μm) and the short-wave IR band (0.9–2.5 μm). The passive IR spectrum consists of the Mid-Wave IR (MWIR) (3–5 μm), and long-wave IR (LWIR)] (7–14 μm) bands. In the passive IR spectrum, heat is exuded from the target, in this particular case the subject's face, and detected by sensors during acquisition. Passive IR sensors are beneficial in challenging conditions, and provide the added benefit of being obscure and difficult to detect. The combination of passive IR sensors with other IR sensors (e.g., IR) can help improve FR accuracy where illumination may be an uncertainty.

This material is based upon work supported by the Center for Identification Technology Research and the National Science Foundation under Grant No. 1066197.

✉ Nnamdi Osia
nosia@mix.wvu.edu
Thirimachos Bourlai
thbourlai@mail.wvu.edu

¹ MILab LCSEE, West Virginia University, 395 Evansdale Drive, Morgantown, WV 26506, USA

1.1 Goals and contributions

Through this work we are able to achieve four goals and contribute to the challenge of FR in heterogeneous environments. First, a database composed of two separate datasets of frontal face images consistent of paired VIS-MWIR and VIS-LWIR face images (using different methods for pre-processing and registration prior to synthesis), are assembled. There are: (1) face detection, (2) CSU geometric normalization, and (3) our recommended geometric normalization method. Through the use of two databases collected under different registration techniques (e.g., captured in two different spectrum at the same time), we are able to quantitatively examine the importance of co-registration. The generated datasets reflect the challenges of face alignment given our patch-based approach for cross-spectral matching. Another such challenge is the optimal placement of the synthesized dataset prior to matching (i.e., better used as the gallery or probe set?). Second, we formulate a image synthesis framework and propose post-synthesis restoration methodology. The restoration approach helps demonstrate the improvement of face recognition accuracy for practical scenarios (e.g., where the gallery image is not synthesized). Third, we explore gender-based filtering (face images are tagged based on their gender) in order to increase FR accuracy, instead of matching all images in the gallery to all images in the probe set. Finally, by conducting an extensive experimental study, we establish that it is feasible to match FR images acquired using the passive IR sensors to visible FR images, and vice-versa, with promising results. Our results are compared to a baseline commercial matcher, Colorado State University's academic matchers, and other original texture-based face matchers.

1.2 Paper organization

The remainder of this work can be partitioned as follows. Sections 2, 3, 4, 5, 6, and 7 describe related work, face image synthesis, post-processing (image restoration and denoising), database and methodological steps, and assessment of our approach. We close out the paper with a conclusion and future works.

2 Related works

2.1 Heterogeneous FR

Tang et al. pioneered heterogeneous FR work with a number of approaches to transform a sketch into a visible image (or vice-versa) [19,20,34,39]. A number of approaches have been researched in order to address various challenges of heterogeneous FR matching scenarios. Aside from the generative transformation-based approaches, recent research in

heterogeneous FR utilize approaches that are discriminative feature-based [12–14,16,18,44], and have shown good accuracies for face matching in both the sketch-focused and NIR-based domains. Sarfraz et al. [28] use deep learning methods to benchmark the Carl thermal-visible dataset (NVESD) where there are changing activity levels and variations in subject-to-camera distance, and illumination. Other implementations, use nonlinear dimensionality reduction, manifold learning, and photometric normalization for optimal feature discrimination based on the spectrum of operation.

2.2 Image synthesis

The upside of synthesis-based methods is that once conversion has been completed, existing FR algorithms can be used for matching. We review three types of approaches for image synthesis: (i) face synthesis analysis; (ii) subspace methods; (iii) 3D-based approaches.

- *Face synthesis analysis* Li et al. [17] propose a stereoscopic synthesis method that produces frontal face images based on two different poses of face images that are co-captured. In [38] face images are transformed from one type to another using face analogy software and then subsequently synthesized query images are matched against gallery images. Zhang et al. [45] developed a face synthesis approach where corresponding sparse coefficients of visible and NIR images are assumed to be alike through learning pairs of an over-complete dictionary.
- *Subspace methods* In [22] the authors augment a challenging database consistent of just one sample per subject by synthesizing new face samples of various degrees using edge-based information. Yi et al. [41] and Dou et al. [10] utilized canonical correlation analysis (CCA) to learn the relationship between face pairs using 9 out of 10 samples from each subject for the training algorithm, and the remaining sample for conversion. Recently, Lei and Li [15] suggested solving the same problem via a low dimensional representation for each face, using a discriminative graph embedding method.
- *3D-based methods* Video can be used to extract 3D features instead of utilizing a 2D face image. Ansari et al. [1] created a database of 3D textured face models composed of 114 subjects using stereo images and a generic face mesh model for 3D FR application. In [21] a 3D generic face model is aligned with each frontal face image.

3 Face image synthesis

An example of our image synthesis workflow is provided in Fig. 1. Please note that unlike other heterogeneous

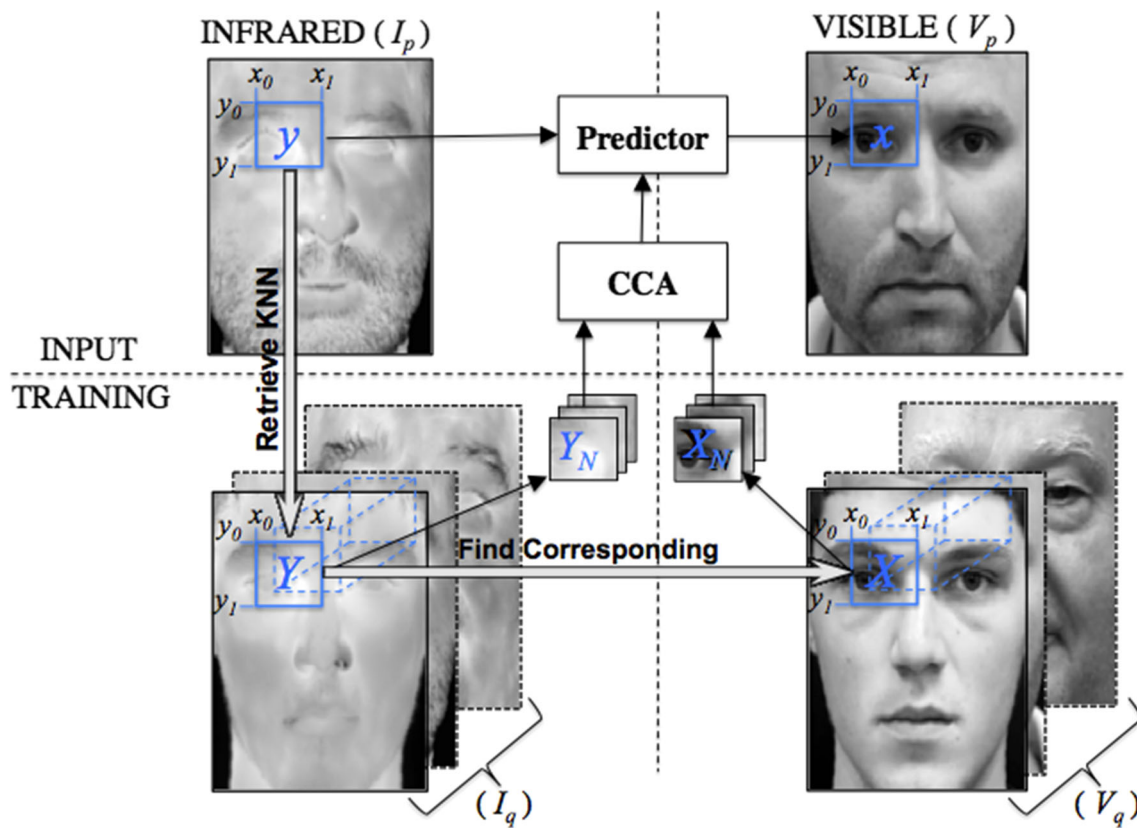


Fig. 1 Flow chart of image synthesis

thermal-visible matching approaches, we use only the facial information (after face detection and normalization) for synthesis, restoration and matching. We do not use the entire thermal head signature that includes more features that may result in enhanced accuracy as for example in [30].

3.1 Canonical correlation analysis

Through the use of two random variables with zero-mean \mathbf{x} , a $p \times l$ vector, and \mathbf{y} , a $q \times l$ vector, CCA finds the 1st pair of directions \mathbf{w}_1 and \mathbf{v}_1 with maximum correlation between the projections $x = \mathbf{w}_1^T \mathbf{x}$ and $y = \mathbf{v}_1^T \mathbf{y}$, $\max \rho(\mathbf{w}_1^T \mathbf{x}, \mathbf{v}_1^T \mathbf{y})$, s.t. $\text{Var}(\mathbf{w}_1^T \mathbf{x} = 1)$ and $\text{Var}(\mathbf{v}_1^T \mathbf{y} = 1)$, where the correlation coefficient is ρ , the variables x and y are known as the first canonical variates, and the \mathbf{w}_1 and \mathbf{v}_1 represents the initial correlation direction vector. CCA finds k th pair of directions \mathbf{w}_k and \mathbf{v}_k which satisfies: (1) $\mathbf{w}_k^T \mathbf{x}$ and $\mathbf{v}_k^T \mathbf{y}$ are not correlated with the previous $k-1$ canonical variates; (2) the correlation between $\mathbf{w}_k^T \mathbf{x}$ and $\mathbf{v}_k^T \mathbf{y}$ is optimized under the constraints $\text{Var}(\mathbf{w}_1^T \mathbf{x} = 1)$ and $\text{Var}(\mathbf{v}_1^T \mathbf{y} = 1)$. Then $\mathbf{w}_k^T \mathbf{x}$ and $\mathbf{v}_k^T \mathbf{y}$ are called the k th canonical variates, and \mathbf{w}_k and \mathbf{v}_k are the k th correlation direction vector, $k \leq \min(p, q)$. The solution for the correlation of coefficients and directions is not different from the generalized eigenvalue problem seen here,

$$(\Sigma_{xy} \Sigma_{yy}^{-1} \Sigma_{xy}^T - \rho^2 \Sigma_{xx}) \mathbf{w} = 0, \quad (1)$$

$$(\Sigma_{xy}^T \Sigma_{xx}^{-1} \Sigma_{xy} - \rho^2 \Sigma_{yy}) \mathbf{v} = 0, \quad (2)$$

where Σ_{xx} and Σ_{yy} are the self-correlation while the Σ_{xy} and Σ_{yx} are the co-correlation matrices respectively. Through CCA, the correlation of the two data sets are prioritized, unlike PCA, which is designed to minimize the reconstruction error. Generally speaking, a few projections (canonical variates) are not adequate to recover the original data well enough, so there is no guarantee that the directions discovered through CCA cover the main variance of the paired data. In addition to the recovery problem, the overfitting problem should be accounted and taken care of as well. If a small amount of noise is present in the data, CCA is so sensitive it might produce a good result to maximize the correlations between the extracted features, but the features may likely model the noise rather than the relevant information in the input data. In this work we use a method called regularized CCA [23]. This approach has proven to overcome the overfitting problem by adding a multiple of the identity matrix $\lambda \mathbf{I}$ to the co-variance matrix Σ_{xx} and Σ_{yy} .

3.2 Feature extraction using CCA

Local features are extracted, instead of features that are holistic, because the latter features seem to fail capturing localized

characteristics and facial traits. The datasets used in training CCA consists of paired VIS and IR images. The images are divided into patches that overlap by the same amount at each position, where there exists a set of patch pairs for CCA learning. CCA locates directional pairs $\mathbf{W}^{(i)} = [\mathbf{w}_1, \mathbf{w}_2, \dots, \mathbf{w}_k]$ and $\mathbf{V}^{(i)} = [\mathbf{v}_1, \mathbf{v}_2, \dots, \mathbf{v}_k]$ for VIS and IR patches, respectively, where the superscript (i) represents the index of the patch (or the location of the patch within the face image). Each column of \mathbf{W} or \mathbf{V} is a directionary vector, which is unitary, but between different columns it is not orthogonal. For example, if we take a VIS patch \mathbf{p} (which can be vectorized as a column) at position i , we are able to extract the CCA feature of the patch \mathbf{p} , using $\mathbf{f} = \mathbf{W}^{(i)T} \mathbf{p}$, where \mathbf{f} is the feature vector belonging to the patch. For each patch and each position at each patch, we are able to acquire CCA projections using our pre-processed training database face images. Projection onto the proper directions is used to extract features, then at each patch location i we get the VIS $\mathbf{O}_v^i = \{\mathbf{f}_{v,j}^i\}$ and IR training sets $\mathbf{O}_{ir}^i = \{\mathbf{f}_{ir,j}^i\}$ respectively.

3.3 Reconstruction using features

In our reconstruction phase that occurs during testing, we use explicitly learned LLE weights in conjunction with our training data to reconstruct the patch and preserve the global manifold structure. Reconstructing the original patch \mathbf{p} through the vectorized feature \mathbf{f} is an arduous task. We are unable to recover the patch by $\mathbf{p} = \mathbf{W}\mathbf{f}$ as we do in PCA because \mathbf{W} is not orthogonal. However, the original patch can be obtained by solving the least squares problem below,

$$\mathbf{p} = \arg_p \min \|\mathbf{W}^T \mathbf{p} - \mathbf{f}\|_2^2, \quad (3)$$

or to add an energy constraint,

$$\mathbf{p} = \arg_p \min \|\mathbf{W}^T \mathbf{p} - \mathbf{f}\|_2^2 + \|\mathbf{p}\|_2^2. \quad (4)$$

The least squares problem can be solved effectively using the scaled conjugate gradient method. In order for the above reconstruction method to be feasible, the feature vector \mathbf{f} has to contain enough information about the original patch. The original patch can be recovered using LLE [27] when fewer features, represented as canonical variates, can be extracted. The assumption that localized geometries pertaining to the manifold of the feature space and that of the patch space are similar, is taken into consideration (see [11]). The patch from the image to be converted and its corresponding features have similar reconstruction coefficients. If $\mathbf{p}_1, \mathbf{p}_2, \dots, \mathbf{p}_k$ are the patches whose features $\mathbf{f}_1, \mathbf{f}_2, \dots, \mathbf{f}_k$ are \mathbf{f} 's k nearest neighbors, and \mathbf{f} is able to be recovered using neighboring features with $\mathbf{f} = \mathbf{F}\mathbf{w}$, where $\mathbf{F} = [\mathbf{f}_1, \mathbf{f}_2, \dots, \mathbf{f}_k]$, $\mathbf{w} = [\mathbf{w}_1, \mathbf{w}_2, \dots, \mathbf{w}_k]^T$, we can reconstruct the original patch using $\mathbf{p} = \mathbf{P}\mathbf{w}$, where $\mathbf{P} = [\mathbf{p}_1, \mathbf{p}_2, \dots, \mathbf{p}_k]$. Using a

probe IR image, we divide it into smaller patches, and obtain the feature vector \mathbf{f}_{ir} of every patch. When we infer the corresponding VIS feature vector \mathbf{f}_v , the VIS patch can be obtained using $\mathbf{p} = \mathbf{P}\mathbf{w}$ for reconstruction and then the patches will be combined into a VIS facial image.

4 Face image restoration and denoising

Unwanted noise is introduced into the image through the image synthesis process (see Fig. 3). Therefore, image denoising [24] is considered as a worthy post-synthesis step that could help improve FR accuracy. Simple image filtering is not ideal for recovering useful image content because it can remove important frequency components in the pipeline. To help alleviate the challenge of effective removal of noise and subsequent image restoration, linear denoising (e.g., filtering), and nonlinear denoising (e.g., thresholding) can be combined to alleviate the noise introduced during image synthesis.

5 Face recognition

5.1 Database

A total of two different datasets were utilized for our experiments. One of the datasets was collected and assembled for our experiments in our laboratory. Each dataset consists of only full frontal face images with a neutral facial expression for every subject. Three different methods are performed for pre-processing across all subjects in both datasets prior to synthesis. A total of about 128 subjects were used in the construction of our database. Each subject had 4 samples that were used for our matching and synthesis experiments (see Table 1).

- **WVU** The (1) VIS-MWIR subset consists of 308 bitmap images (154 for probe and 154 for gallery) with four sequential images in time per subject (77 subjects). Visible images for this database were extracted from videos captured in our laboratory, using a Canon EOS 5D Mark II camera, where a full image contained the subject's complete head and shoulders. This digital SLR camera produces ultra-high-resolution RGB color images or videos, with a resolution of 1920×1080 pixels. The face images are obtained by obtained from the movie files in JPEG format. The MWIR images for this database were extracted from videos captured in our laboratory, using a FLIR SC8000 MWIR camera, where a full image contained the subject's complete head and shoulders. The infrared camera produces high-definition thermal videos, with a resolution of 1024×1024 . The (2) VIS-LWIR

Table 1 Brief description of the database used in empirical evaluation of the proposed approach

Database			
	Datasets	Description	Number of (subjects/samples per subject)
WVU	1. Visible & MWIR	Controlled face (Canon Mark II and FLIR SC800)	77/4
	2. Visible & LWIR	Controlled face (Canon Mark II and FLIR SC600)	78/4
NVESD	1. Visible & MWIR	Controlled face (Basler Scout sc640 and DRS MWIR)	50/4
	2. Visible & LWIR	Controlled face (Basler Scout sc640 and DRS LWIR)	50/4

This database is constructed using three different methods for pre-processing face images prior to image synthesis methods for WVU dataset and NVESD dataset

subset consists of 312 bitmap images (156 for probe and 156 for gallery) with four sequential images in time per subject (78 subjects). Visible images for this subset were extracted from videos captured using the aforementioned Canon EOS 5D Mark II camera. The LWIR images for this subset were extracted from videos captured in our laboratory, using a FLIR SC600 LWIR camera, where a full image contained the subject's complete head and shoulders. The science-grade infrared camera produces high-resolution LWIR images or videos, with a resolution of 640×480 pixels. The first 2 samples were utilized as gallery images, while the remaining 2 samples were the probe images. It is noteworthy that images between sensor pairs were not captured simultaneously or co-registered (e.g., captured in both bands at the same time), making our database more challenging given our patch-based approach.

- *NVESD* The NVESD dataset [7] was acquired as a joint effort between the Night Vision Electronic Sensors Directorate of the U.S. Army Communications-Electronics Research, Development and Engineering Center (CERDEC), and the U.S. Army Research Laboratory (ARL). The portion of NVESD dataset examined two experimental conditions: vigorous exercise in the form of a fast paced walk and subject-to-camera range (1, 2, and 4 m). A group of 25 subjects were imaged before and after exercise at each of the three ranges. Another group of 25 subjects were at rest and imaged at each of the three ranges. All 50 subjects were used to create the dataset; however, only a subject-to-camera range of 1 and 2 m was used for our dataset. For the (1) VIS-MWIR subset, visible images were captured using the Basler Scout GigE Vision Sensor sc640-74gm equipped with Sigma 24 mm f/1.8 EX DG Aspherical Macro Large Aperture Wide Angle Lenses. The visible sensor was used to acquire 8-bit grayscale facial images and was connected to GigE via a Netgear router connected to the collection PC. The sensor has pixel pitches of $10 \mu\text{m}$ and spectral responses of 400–1000 nm, with a peak at 500 nm. The images were acquired for 15 s per capture at a resolution of 640×480 at 30 Hz. Software known as JAI Camera Control Tool was

used to obtain the images and store them in raw, uncompressed AVI and TIFF formats. The MWIR face images were acquired using a DRS sensor. The DRS sensor was used to acquire 16-bit (12-bit) grayscale facial images in each band. The MWIR sensor has a pixel pitch of $12 \mu\text{m}$ and a spectral response of 3–5 μm . The same aforementioned Basler Scout sc6470 camera was used to acquire visible images for the (2) VIS-LWIR subset. The LWIR face images were acquired using a DRS sensor. The DRS sensor was used to acquire 16-bit (12-bit) grayscale facial images in each band. The LWIR sensor has a pixel pitch of $15 \mu\text{m}$ and a spectral response of 8–12 μm . Images were acquired for 15 s per capture at a resolution of 640×480 at 30 Hz. AutoIt software was used to acquire the images and store them in the .raw format. Each acquisition lasted 15 s at 30 frames per second for each camera, with all the sensors started and stopped almost simultaneously (subject to slight offsets because of human reaction time). To form a set of gallery and probe images for face recognition, a frame was extracted at the 1 and 14 s marks for each video. The first 2 samples of the gallery and probe sets, respectively, were constructed using the still frames of a 1 and 2 m subject-to-camera range. Images between sensor pairs were captured simultaneously.

5.2 Methodological steps

The entire overview of this framework is illustrated in Fig. 2. The pertinent stages of the methodology proposed in this work are described below:

1. *Pre-processing* Our proposed approach is patch-based; therefore, it is important that the correct corresponding patches overlap as precisely as possible in both spectra. We experiment with three different face image pre-processing techniques, all discussed in detail below. The metric we use for performance evaluation is rank-1 identification accuracy (CMC). The left and right eye coordinates are manually annotated on the raw images prior to pre-processing. Samples of the face images after pre-processing can be seen in Fig. 3.

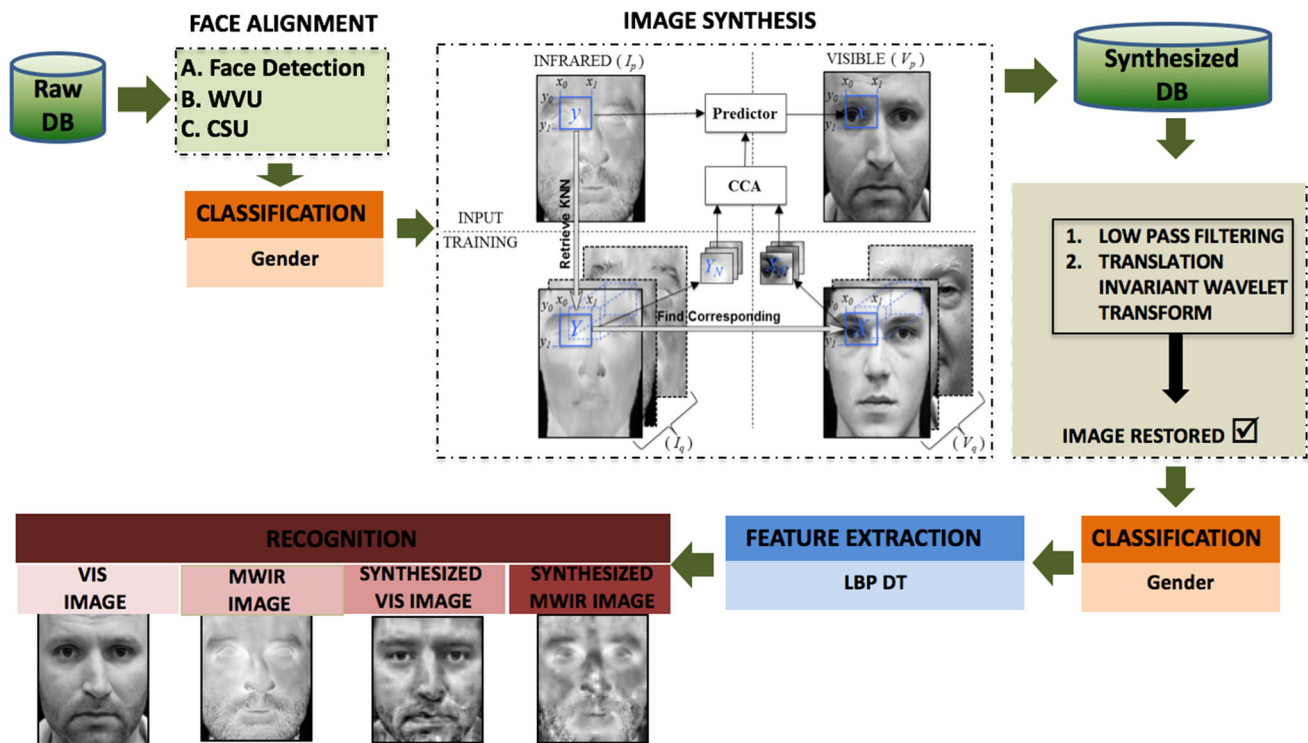


Fig. 2 Schematic of the proposed FR methodology which consists of normalization, synthesis, restoration, and matching

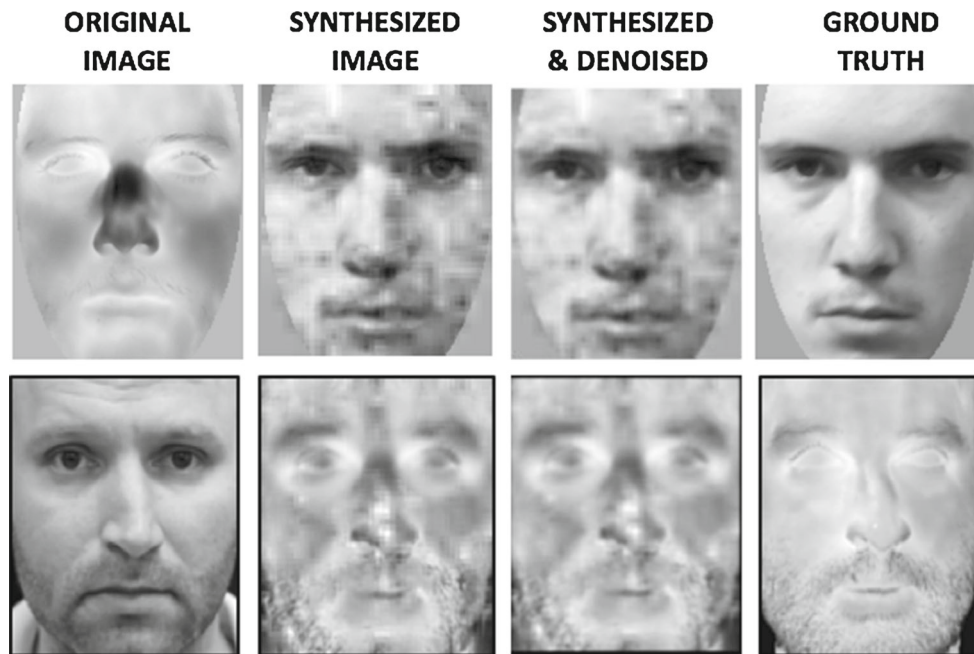


Fig. 3 Example original, synthesized, synthesized and denoised and ground truth images from two separate subjects. The subject on *top row* (MWIR to VIS) was normalized using CSU normalization, while the subject on the *bottom row* (VIS to MWIR) was normalized using our proposed normalization technique

– *Face detection* For the visible spectrum of our database, Viola & Jones face detection algorithm [37] is used to determine the rectangular overlay or boundary around the face. This algorithm has been

regarded to perform efficiently on facial images captured in the visible spectrum, but additional training is necessary for the passive IR band. However, there were still several limitations when Viola & Jones

is applied to the passive IR band of our database, due to the lack of training data (not many available and the operational cost to collect more with both our cameras was prohibited). To compensate, blob detection-based approach is applied in our passive infrared band images, resulting in 85% better detection accuracy than Viola & Jones (whose haar cascades are trained specifically for visible data).

- *CSU normalization* Colorado State University's (CSU) Face Identification Evaluation System [3] FR software is first utilized for pre-processing. The normalization is a spatial transformation, which utilizes the left and right eyes as control points. Shapes in the original image are unchanged, but the image is distorted by a combination of translation, rotation, and scaling. After geometric normalization, the image is cropped using an elliptical mask so that only the face from the forehead to the chin and cheek to cheek can be seen.
- *Normalization (proposed)* A standard interocular distance is set and the eye locations are centered and aligned onto a single horizontal plane and resized to fit the desired distance. Each face image was geometrically normalized based on the manually found locations to have an interocular distance of 60 pixels with a resolution of 111×121 pixels. There is no elliptical mask applied in our approach, in contrast to the CSU normalization software.

2. *Image synthesis and restoration* The formulated image synthesis methodology is combination of manifold learning and nonlinear dimensionality reduction. We utilize the leave-one-out method during synthesis, where the sample left out of the training set is used for conversion from one spectrum to another. Through the image synthesis algorithm, we are able to convert the datasets described and create their synthesized versions. After the synthesized data is created, it is later used for identity authentication. We restore the synthesized images from the previous step using a combination of linear denoising and thresholding. Noniterative denoising methods are a possible solution for the noise problem through numerical calculations that are explicitly solved. Noniterative methods are usually easier to implement and are not computationally complex.
3. *Face recognition systems* We utilize the *Local Binary Patterns* (LBP) method [33] for FR due to its previous use and success with the cross-spectral face recognition problem [5]. The LBP operator is an efficient, nonparametric, and unifying approach to traditional divergent models for analyzing texture that are statistical and structural based. Occurrences of different binary patterns are then counted up using a histogram. The cumulative match characteris-

tic (CMC) curve is used to measure the identification accuracy of the system. With this metric, the ranking potential of the system can be measured, showing the $1 : m$ identification performance.

5.3 Gender classification

Gender-based cohort classification is achieved using an approach that detects Histogram of oriented gradient (HOG) features [8] and a support vector machine (SVM) classifier [25]. In order to train the classifiers, it is important that vectorized HOG features are extracted from the images used for training. The extracted HOG feature that is vectorized should be capable of encoding an precise amount of information pertaining to the subject. SVM is a kernel based method and has mostly been used for two class classification [25]. Through the use of nonlinear mapping, kernel algorithms are capable of mapping data from a original space into a higher dimensional feature space. The downside of this is that in high dimensional spaces, the curse of dimensionality is evident, although there exists a workaround for finding the scalar products in the feature space. When considering two feature space vectors, calculation of the scalar product can be done explicitly with the help of kernel functions.

6 Empirical evaluation

The experimental scenarios we evaluate in this work are as follows: (1) baseline experiments; (2) optimization of image synthesis; (3) post-synthesis image restoration w.r.t. FR accuracy; (4) automatic classification experiments; and (5) identification performance after gender-based filtering. After optimizing our selected matcher for the given problem (e.g., LBP/LTP), the distance transform (DT) appears to be a more consistent method in achieving higher FR accuracy. When comparing selected matchers (e.g., LBP vs LTP), LBP holds a slight edge over LTP in many scenarios. For our selected texture-based matcher (e.g., LBP DT), we evaluate the challenge of image alignment using varied pre-processing within our proposed synthesis approach during experimentation. We trained our synthesis and classification algorithms using a leave-one-out approach, i.e., take one image sample out of the training dataset and use that sample as test image for synthesis (the IR image as the input and the VIS image as the ground truth, and vice-versa); the remaining samples of the subject are used for training within our system.

6.1 Baseline experiments

We employ a set of baseline experiments (cross-spectral face matching) by using commercial and academic based software: (1) Commercial-of-the-shelf (COTS) identity software

Table 2 Baseline Rank-1 FR results (%) for VIS-MWIR and VIS-LWIR face matching experiments

WVU Rank-1 raw baseline (CSU) FR accuracy (%)				
Methodology	VIS-LWIR	LWIR-VIS	VIS-MWIR	MWIR-VIS
L1 Systems (G8)	62.82	61.54	40.26	37.01
Bayesian MAP	11.54	11.54	7.69	5.13
Bayesian ML	11.54	10.26	7.69	5.13
LDA Euclidean	11.54	19.23	7.69	8.97
LDA IdaSoft	14.10	19.23	8.97	6.41
PCA Euclidean	8.97	5.13	7.69	6.41
PCA MahCosine	15.38	15.38	5.13	7.69

The bold values are the maximum FR accuracies across gallery and probe

Table 3 NVESD dataset baseline Rank-1 FR results (%) for VIS-MWIR and VIS-LWIR CSU face-matching experiments

NVESD Rank-1 raw baseline (CSU) FR accuracy (%)				
Methodology	VIS-LWIR	LWIR-VIS	VIS-MWIR	MWIR-VIS
Bayesian MAP	38.00	10.00	20.00	10.00
Bayesian ML	32.00	12.00	18.00	12.00
LDA Euclidean	32.00	30.00	14.00	10.00
LDA IdaSoft	32.00	30.00	16.00	16.00
PCA Euclidean	20.00	14.00	4.00	4.00
PCA MahCosine	32.00	28.00	16.00	16.00

The bold values are the maximum FR accuracies across gallery and probe

tools (G8) provided by MorphoTrust (formerly L1); (2) Face Identification Evaluation System which contains standard training-based face recognition methods developed by the CSU [3], including PCA [9,31,36], PCA + LDA [2], the *Bayesian Intrapersonal/Extra-personal Classifier* (BIC) using either the maximum likelihood (ML) or the Maximum a posteriori (MAP) hypothesis [35]. Distance metrics such as Euclidean Distance (EU) are used by both PCA and LDA, which result in the ordinary or standard distance between two feature vectors (PCA EU + LDA EU); (3) *Local Binary Pattern* (LBP) method [26], as aforementioned in the methodological steps for our face recognition pipeline.

Utilizing a commercial matcher (G8), the rank-1 identification rate achieved is 40.26% for the WVU VIS-MWIR dataset and 62.82% for the WVU VIS-LWIR dataset. For CSU academic matcher, the maximum rank-1 identification rate recorded is 19.23% for WVU VIS-LWIR and 8.97% for WVU VIS-MWIR using the LDA algorithm. For our NVESD dataset, the Bayesian algorithm had the best results with a rank-1 identification rate of 38.00% for NVESD VIS-LWIR and 20.00% for NVESD VIS-MWIR combination. With regard to the LBP DT matcher, the rank-1 identification rate achieved is 5.29% for the WVU VIS-MWIR spectrum. For the WVU VIS-LWIR spectrum, the rank-1 identification rate achieved for the LBP DT matcher is 5.19%. When evaluating the LBP DT matcher on the NVESD VIS-MWIR dataset, a rank-1 of 20.00% was achieved. For the LBP DT matcher on the NVESD VIS-LWIR dataset, a rank-1 of 14.00% was recorded. The baseline results using COTS G8

Table 4 WVU dataset baseline Rank-1 FR results (%) for pre-processed VIS-MWIR and VIS-LWIR face-matching experiments (LBP DT)

Gallery	Probe	Baseline Rank-1 FR accuracy (%)		
		Face detect	Proposed (WVU)	CSU
VIS	MWIR	4.55	3.90	5.19
MWIR	VIS	3.90	3.90	1.95
VIS	LWIR	2.56	4.49	4.49
LWIR	VIS	3.85	5.13	1.92

The bold values are the maximum FR accuracies across gallery and probe

for the WVU dataset is shown in Table 2. We are unable to provide results for COTS G8 algorithm with the NVESD dataset for this work. For training-based academic matcher CSU, the baseline results for both the WVU and NVESD datasets can be seen in Tables 2 and 3, respectively. The baseline results using texture-based LBP DT matcher can be seen in Tables 4, and 5 for both datasets.

6.2 Image synthesis experiments

There are several parameters to be chosen in our proposed synthesis algorithm, such as the size of patches, the number of canonical variates k (the dimensionality of feature vector) we take for every patch, and the number of the neighbors we use to train the canonical directions. Generally speak-

Table 5 NVESD dataset baseline Rank-1 FR results (%) for pre-processed VIS-MWIR and VIS-LWIR face-matching experiments (LBP DT)

Gallery	Probe	Baseline Rank-1 FR accuracy (%)		
		Face detect	Proposed (WVU)	CSU
VIS	MWIR	6.00	12.00	6.00
MWIR	VIS	16.00	20.00	20.00
VIS	LWIR	4.00	8.00	6.00
LWIR	VIS	14.00	14.00	16.00

The bold values are the maximum FR accuracies across gallery and probe

ing, the correlation between pairs of IR and VIS patches of a smaller size is weaker, so the inference is less reasonable. While the larger the patch size, makes the correlation stronger, more canonical variates are needed to represent the patch, which makes training samples much sparser in the feature space. The size of all the images in our database despite pre-processing methodology is 320×256 during the synthesis step, and we choose a patch size of 9×9 with 3-px overlapping. Since the projections (features) onto the former pairs of direction have stronger correlations, choosing fewer features makes the inference more robust, while choosing more features gives a more precise adaptation of the original

patch. Similarly, when we choose a larger number of neighbors, K , there are more samples, which makes the algorithm more robust but computationally expensive. We choose 5 features and 100 neighbors for LLE. Once we have converted a spectrum from VIS to MWIR, MWIR to VIS, VIS to LWIR, or LWIR to VIS, our heterogeneous cross-spectral matching problem can now be considered to be an homogenous intra-spectral matching problem again. CLAHE normalization is applied to both sets of gallery and probe images after synthesis and prior to matching. Although not practical, our matching experiments after synthesis are tested using the synthesized images as both gallery and probe set, for each spectrum.

With respect to the WVU dataset, we achieve a maximum rank-1 identification rate of 85.06% when using LBP DT matcher after synthesis for WVU VIS-MWIR and a maximum rank-1 identification rate of 79.49% for the WVU VIS-LWIR spectrums, using WVU normalization. For the NVESD dataset, we achieve a maximum rank-1 identification rate of 98.00% when using LBP DT matcher after synthesis for NVESD VIS-MWIR and a rank-1 identification rate of 100.00% for the NVESD VIS-LWIR spectrums, using CSU normalization. The synthesis results can be seen in Tables 6 and 7 for the synthesized WVU and NVESD datasets, respectively.

Table 6 Rank-1 FR results (%) for synthesized WVU VIS-MWIR and VIS-LWIR datasets using selected matcher (LBP DT)

WVU synthesized Rank-1 FR accuracy (%) LBP DT				
Gallery	Probe	Face detect	Proposed (WVU)	CSU
VIS	Synth. VIS (MWIR)	19.48	31.17	27.27
Synth. VIS (MWIR)	VIS	68.18	76.62	72.73
MWIR	Synth. MWIR	24.68	35.71	26.62
Synth. MWIR	MWIR	58.44	85.06	76.62
VIS	Synth. VIS (LWIR)	8.97	40.38	27.56
Synth. VIS (LWIR)	VIS	53.85	85.26	66.67
LWIR	Synth. LWIR	26.28	38.46	45.51
Synth. LWIR	LWIR	56.41	79.49	77.56

The bold values are the maximum FR accuracies across gallery and probe

Table 7 Rank-1 FR results (%) for synthesized NVESD VIS-MWIR and VIS-LWIR datasets using selected matcher (LBP DT)

NVESD Synthesized Rank-1 FR accuracy (%) LBP DT				
Gallery	Probe	Face detect	Proposed (WVU)	CSU
VIS	Synth. VIS (MWIR)	50.00	68.00	74.00
Synth. VIS (MWIR)	VIS	86.00	86.00	84.00
MWIR	Synth. MWIR	62.00	56.00	88.00
Synth. MWIR	MWIR	90.00	94.00	98.00
VIS	Synth. VIS (LWIR)	54.00	68.00	76.00
Synth. VIS (LWIR)	VIS	86.00	84.00	78.00
LWIR	Synth. LWIR	80.00	60.00	39.00
Synth. LWIR	LWIR	90.00	96.00	100.00

The bold values are the maximum FR accuracies across gallery and probe

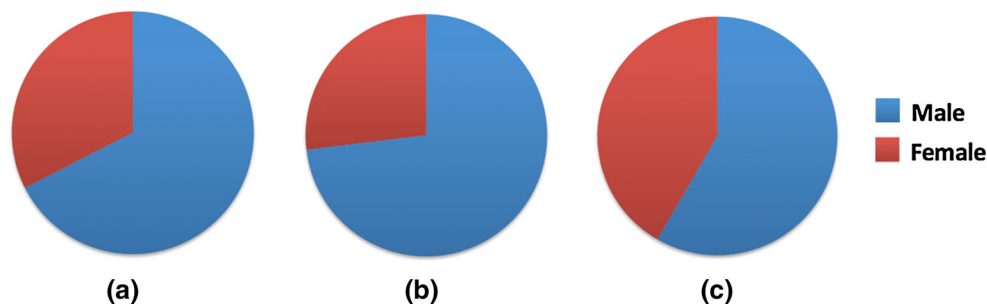


Fig. 4 Gender-based filtering of dataset for **a** WVU VIS-MWIR, **b** WVU VIS-LWIR and **c** NVESD VIS-IR. NVESD VIS-MWIR and NVESD VIS-LWIR both have the same amount of subjects (50) and male-female proportion

Table 8 Rank-1 FR results (%) for restored synthesized WVU VIS-MWIR and VIS-LWIR datasets using selected matcher (LBP DT)

WVU restored synthesized Rank-1 FR accuracy (%) LBP DT				
Gallery	Probe	Face detect	Proposed (WVU)	CSU
VIS	Synth. VIS (MWIR)	21.43	30.52	31.17
Synth. VIS (MWIR)	VIS	68.83	75.32	77.27
MWIR	Synth. MWIR	27.27	42.86	51.30
Synth. MWIR	MWIR	59.74	81.17	76.62
VIS	Synth. VIS (LWIR)	5.13	37.82	23.72
Synth. VIS (LWIR)	VIS	53.85	81.41	79.49
LWIR	Synth. LWIR	25.00	43.59	57.69
Synth. LWIR	LWIR	60.90	78.85	80.77

The bold values are the maximum FR accuracies across gallery and probe

Table 9 Rank-1 FR results (%) for restored synthesized NVESD VIS-MWIR and VIS-LWIR datasets using selected matcher (LBP DT)

NVESD restored synthesized Rank-1 FR accuracy (%) LBP DT				
Gallery	Probe	Face detect	Proposed (WVU)	CSU
VIS	Synth. VIS (MWIR)	66.00	70.00	78.00
Synth. VIS (MWIR)	VIS	86.00	86.00	92.00
MWIR	Synth. MWIR	74.00	67.00	88.00
Synth. MWIR	MWIR	90.00	92.00	98.00
VIS	Synth. VIS (LWIR)	88.00	71.00	76.00
Synth. VIS (LWIR)	VIS	88.00	82.00	90.00
LWIR	Synth. LWIR	82.00	66.00	88.00
Synth. LWIR	LWIR	90.00	92.00	96.00

The bold values are the maximum FR accuracies across gallery and probe

6.3 Image restoration experiments

In this evaluation, we determine the effects of applying a combination of filtering and TI-denoising on synthesized images in order to improve FR accuracy of our datasets under practical scenarios (e.g., gallery images are not synthesized). Both synthesized and ground truth (gallery and/or probe) sets were LP filtered and subsequently denoised. We optimize our proposed image restoration parameters, LP filter type and sigma value threshold used for TI-denoising, using CMC rank-1 accuracy as a metric. First, we apply an LP Filter to minimize distortion due to the subsampling. The type of LP filter used is a boxcar filter with a fixed win-

dow size. Through previous experimentation [6], we found a window size of 3 to be optimal. After we are able to LP filter the image, denoising is carried out using the TI-denoising scheme. The sigma value chosen for TI-denoising appears to be optimal depending on whether we are denoising synthesized images or ground truth images. Synthesized images received a sigma value of 3, while ground truth images were only slightly denoised with a sigma value of .01.

With respect to the WVU dataset, we achieve a maximum rank-1 identification rate of 81.17% when using LBP DT matcher after synthesis for WVU VIS-MWIR and a rank-1 identification rate of 80.77% for the WVU VIS-LWIR spec-

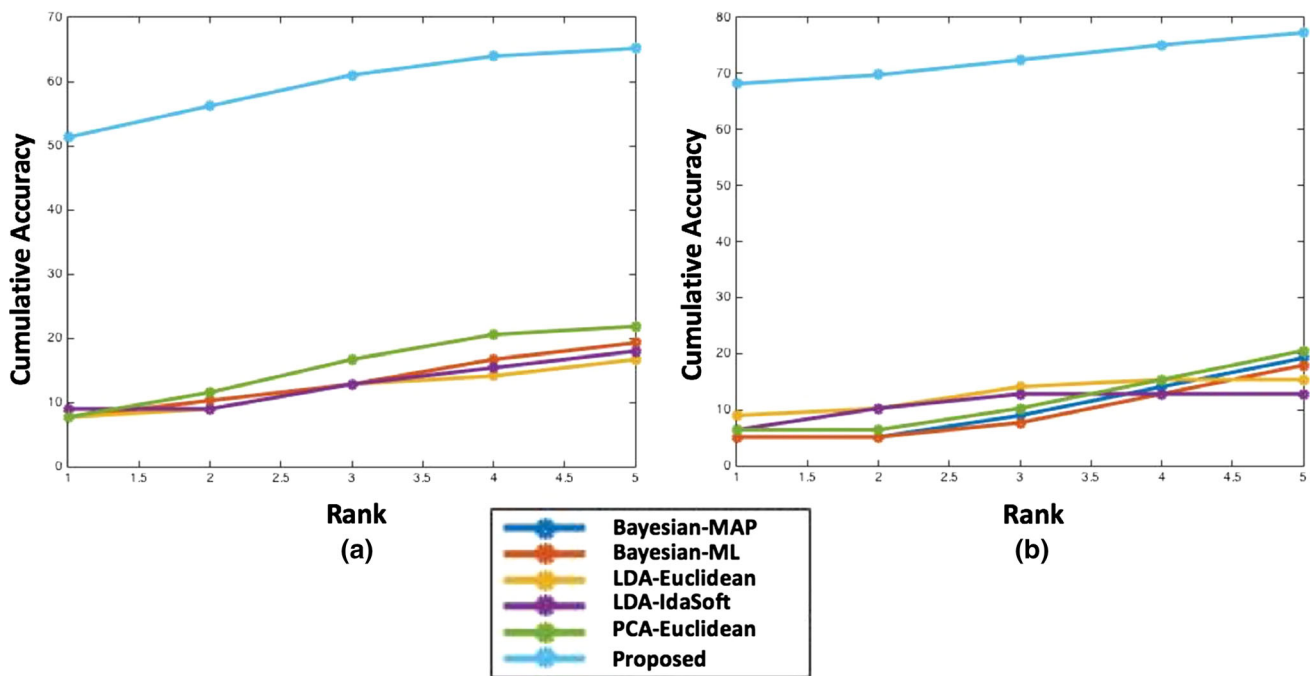


Fig. 5 **a** Identification rates (Rank-1 to Rank-5) for WVU VIS gallery and MWIR probe. **b** Identification rates (Rank-1 to Rank-5) for WVU MWIR gallery and VIS probe

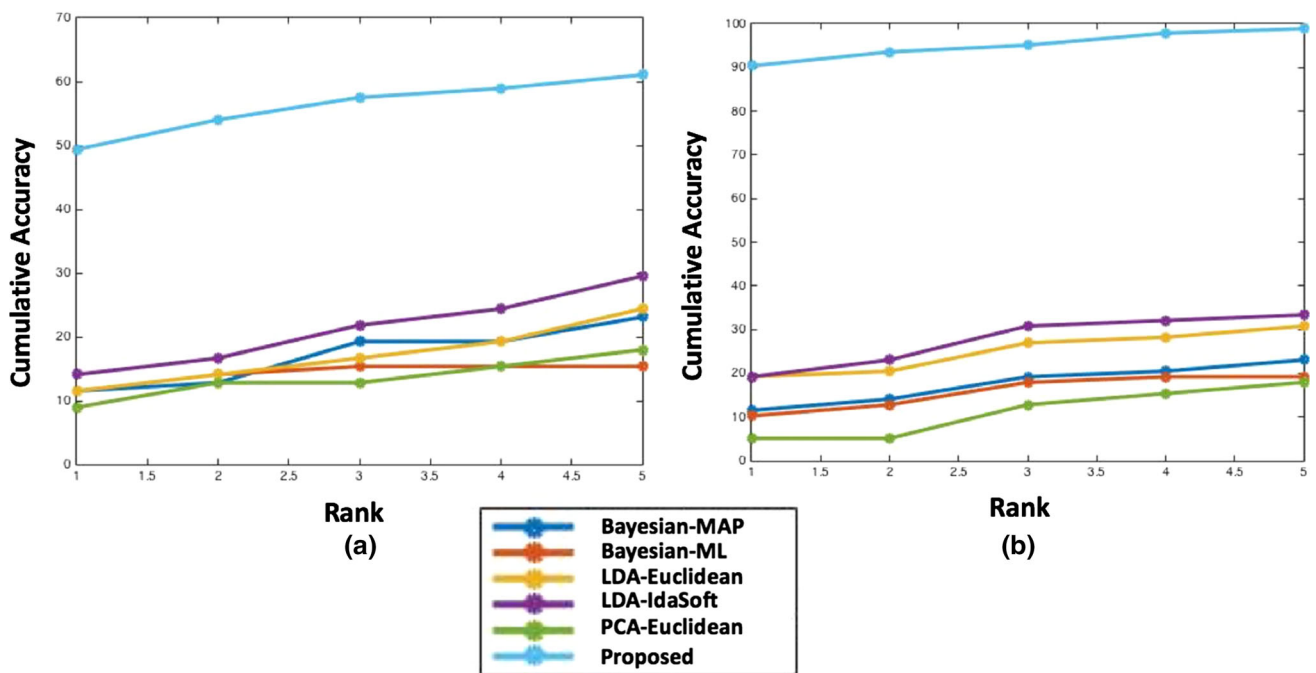


Fig. 6 **a** Identification rates (Rank-1 to Rank-5) for WVU VIS gallery and LWIR probe. **b** Identification rates (Rank-1 to Rank-5) for WVU LWIR gallery and VIS probe

trums, using WVU and CSU normalization, respectively. For the NVESD dataset, we achieve a maximum rank-1 identification rate of 98.00% when using LBP DT matcher for NVESD VIS-MWIR and a rank-1 identification rate of

96.00% for the NVESD VIS-LWIR spectrums, using CSU normalization (Fig. 4).

The results for FR of synthesized images after image restoration can be seen in Tables 8 and 9 for the restored

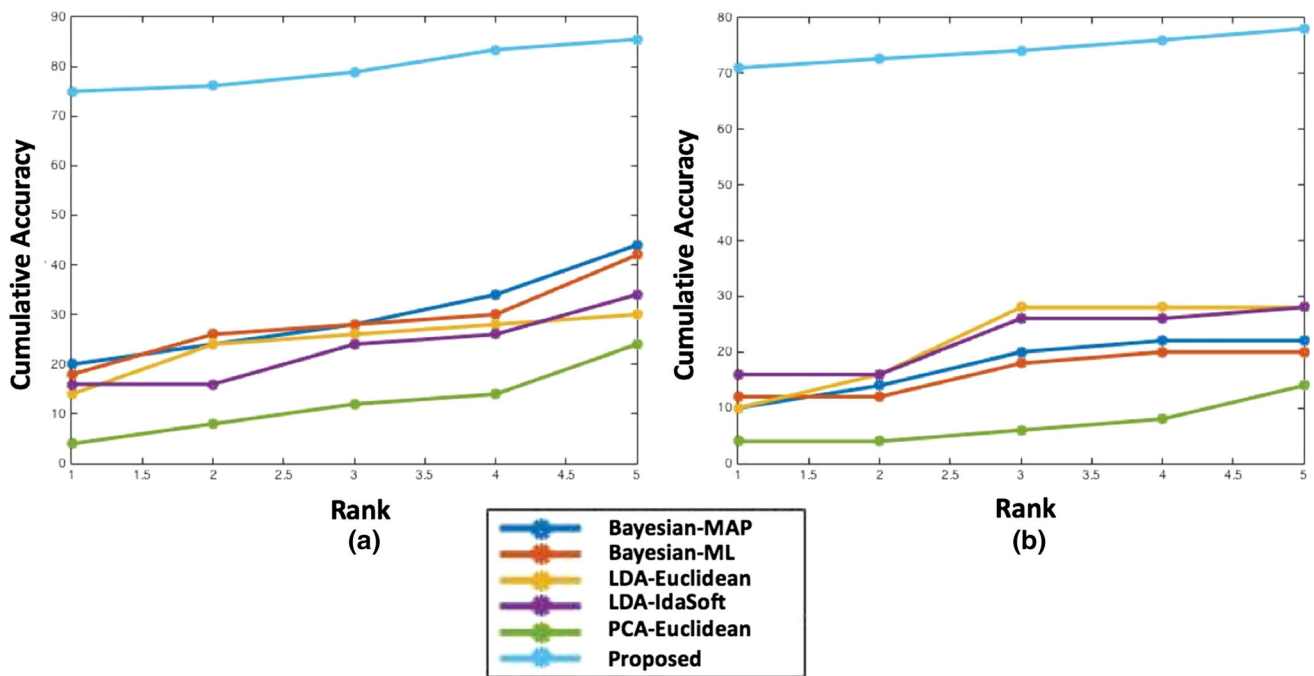


Fig. 7 **a** Identification rates (Rank-1 to Rank-5) for NVESD VIS gallery and MWIR probe. **b** Identification rates (Rank-1 to Rank-5) for NVESD MWIR gallery and VIS probe

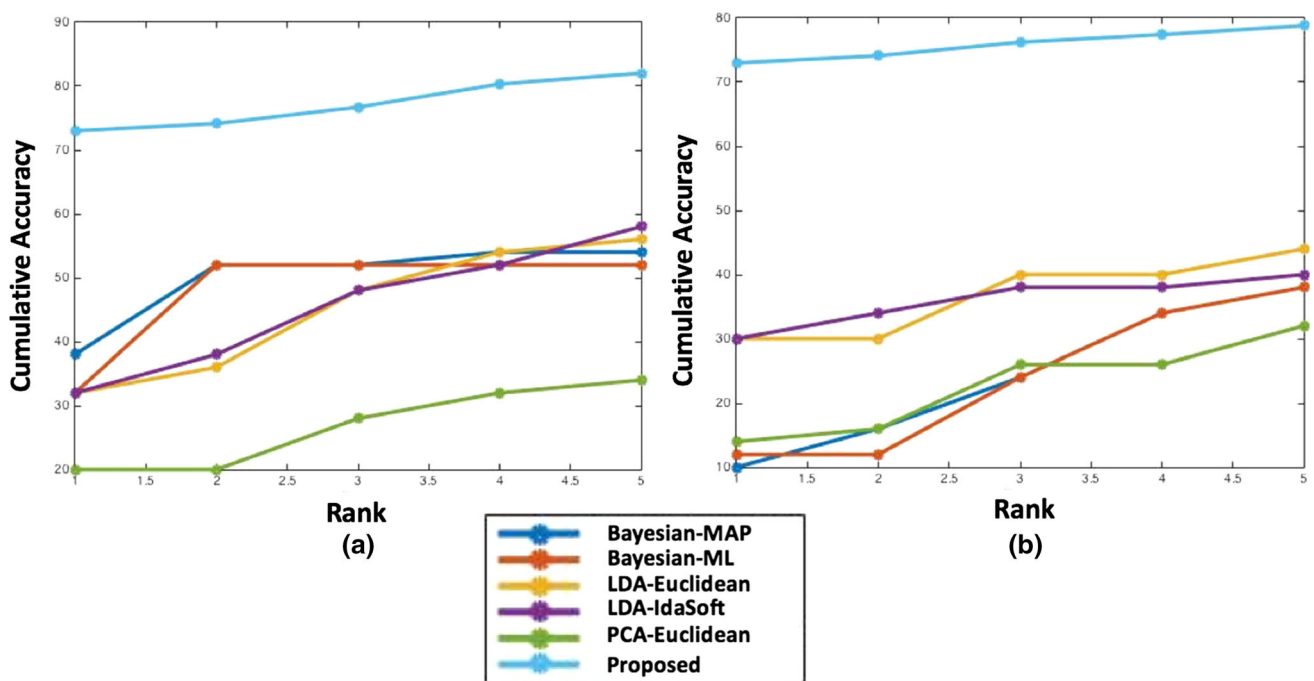


Fig. 8 **a** Identification rates (Rank-1 to Rank-5) for NVESD VIS gallery and LWIR probe. **b** Identification rates (Rank-1 to Rank-5) for NVESD LWIR gallery and VIS probe

WVU and NVESD datasets, respectively. Identification rates (Rank-1 to Rank-5) can be seen for our collected data and proposed methodology (after denoising and image restoration) when compared to classic academic matchers, in Figs. 5, 6, 7 and 8, respectively.

6.4 Gender classification

We achieve gender-based cohort classification scheme using Histogram of oriented gradient (HOG) features [8] and a support vector machine (SVM) classifier [25]. The cell size

Table 10 Rank-1 FR results (%) for cohort filtered and denoised synthesized WVU VIS-MWIR and WVU VIS-LWIR datasets using selected matcher (LBP DT)

WVU cohort filtered denoised synthesized Rank-1 FR accuracy (%) LBP DT				
Gallery	Probe	Face detect	Proposed (WVU)	CSU
VIS	Synth. VIS (MWIR)	42.51	51.30	54.54
Synth. VIS (MWIR)	VIS	84.41	83.77	88.96
MWIR	Synth. MWIR	41.56	68.18	68.83
Synth. MWIR	MWIR	76.62	88.96	85.71
VIS	Synth. VIS (LWIR)	22.44	49.36	41.67
Synth. VIS (LWIR)	VIS	70.51	90.38	88.46
LWIR	Synth. LWIR	43.59	76.28	73.72
Synth. LWIR	LWIR	78.85	90.38	89.10

The bold values are the maximum FR accuracies across gallery and probe

Table 11 Rank-1 FR results (%) for cohort filtered and denoised synthesized NVESD VIS-MWIR and NVESD VIS-LWIR datasets using selected matcher (LBP DT)

NVESD cohort filtered denoised synthesized Rank-1 FR accuracy (%) LBP DT				
Gallery	Probe	Face Detect	Proposed (WVU)	CSU
VIS	Synth. VIS (MWIR)	66.00	75.00	78.00
Synth. VIS (MWIR)	VIS	86.00	86.00	92.00
MWIR	Synth. MWIR	73.00	71.00	96.00
Synth. MWIR	MWIR	90.00	94.00	98.00
VIS	Synth. VIS (LWIR)	63.00	73.00	76.00
Synth. VIS (LWIR)	VIS	88.00	84.00	94.00
LWIR	Synth. LWIR	82.00	73.00	90.00
Synth. LWIR	LWIR	90.00	98.00	96.00

The bold values are the maximum FR accuracies across gallery and probe

that we use to encode the vectorized HOG feature is 4×4 . When training using the leave-one-out scheme, we achieve perfect performance across all pre-processing methods and tested datasets.

6.5 Demographic filtering experiments

The face datasets are filtered by gender and the resulting system performance is evaluated. The objective is to determine whether the hypothesis that face-matching performance improves with a filtering hold, such as gender-based classification. Two different gender subsets, one for each dataset, were used for this experiment. For the gender-based subsets, we split the data into (1) male and (2) female gallery and probe. With respect to the WVU VIS-MWIR gender-based subsets, 52 subjects were male, while 25 subjects were female. For the WVU VIS-LWIR gender-based subsets, 57 subjects were male, while 21 subjects were female. For both the NVESD VIS-MWIR and VIS-LWIR gender-based subsets, 35 subjects were male, while 15 subjects were female. A pie chart representing gender-based distribution for both datasets can be seen in Fig. 4.

With respect to the WVU dataset, we achieve a maximum rank-1 identification rate of 88.96% when using LBP DT matcher after synthesis for WVU VIS-MWIR and a rank-1

identification rate of 90.38% for the WVU VIS-LWIR spectrum, using WVU normalization. For the NVESD dataset, we achieve a maximum rank-1 identification rate of 98.00% when using LBP DT matcher and WVU pre-processing, after synthesis for NVESD VIS-MWIR and a rank-1 identification rate of 96.00% for the NVESD VIS-LWIR spectrum, using WVU and CSU normalization, respectively.

The results for FR of the denoised synthesized images after image restoration and demographic filtering can be seen in Tables 10 and 11 for the synthesized WVU and NVESD datasets, respectively.

7 Conclusions and future work

We study the problem of image synthesis as a means to bridge the informational gap between face images from two different spectral bands. Our study shows that image alignment and co-registration is important in achieving higher FR accuracy for the proposed approach. Experimental results show that recognition accuracy is much higher when the synthesized face image is used as a gallery set, as opposed to the probe set. We believe there is so much difference in rank-1 scores when the synthetic dataset is the gallery vs. the probe because there is more data present in the raw image. When

a synthesized face image is used as the gallery, all information in the synthesized face image should be present in the raw face image of the same subject. However, when the raw face image is the gallery, the synthesized face probe image is likely missing some information that is present in the raw face gallery image. In practical applications, the use of raw face images as the gallery set would be the more realistic scenario. The image restoration step increases the score where we use the slightly denoised raw images as the gallery set, irrespective of the spectral band. However, rank-1 accuracy is decreased when a denoised synthesized image is used as the gallery image. The image restoration step was particularly valuable on the datasets pre-processed using face detection and CSU normalization, excluding matching using VIS gallery and synthesized VIS probe. The image restoration step decreases face recognition accuracy for our proposed geometric normalization pre-processing step. FR software which may be COTS, is difficult to evaluate because it may contain proprietary geometric and photometric normalization, in conjunction with restoration which are unable to be accounted for by the user. This is challenging because when using COTS packages, in addition to ours, a performance drop can be expected when compared to using our proposed approach.

Utilizing our image synthesis approach, we achieve a maximum rank-1 identification rate of 85.06% when using LBP DT matcher after synthesis for WVU VIS-MWIR and a maximum rank-1 identification rate of 79.49% for the WVU VIS-LWIR spectrum, using WVU normalization. For the NVESD dataset, we achieve a maximum rank-1 identification rate of 98.00% when using LBP DT matcher after synthesis for NVESD VIS-MWIR and a rank-1 identification rate of 100.00% for the NVESD VIS-LWIR spectrum, using CSU normalization. After image restoration and denoising of our data, we achieve a maximum rank-1 identification rate of 81.17% when using LBP DT matcher after synthesis for WVU VIS-MWIR and a maximum rank-1 identification rate of 80.77% for the WVU VIS-LWIR spectrum, using WVU and CSU normalization, respectively. For the NVESD dataset, we achieve a maximum rank-1 identification rate of 98.00% when using LBP DT for NVESD VIS-MWIR and a rank-1 identification rate of 96.00% for the NVESD VIS-LWIR spectrum, using CSU normalization. I think it is an important to note that although overall identification rate decreases in scenarios where there is a synthesized gallery set, identification accuracy improves in modes of operation that mimic reality (e.g., when the gallery set is not synthesized). After gender filtering of our synthesized and denoised data, we achieve a maximum rank-1 identification rate of 88.96% when using LBP DT matcher after synthesis for WVU VIS-MWIR and a rank-1 identification rate of 90.38% for the WVU VIS-LWIR spectrum, using WVU normalization. For the NVESD dataset, we achieve a maximum rank-1

identification rate of 98.00% when using LBP DT matcher and WVU pre-processing, after synthesis for NVESD VIS-MWIR and a rank-1 identification rate of 96.00% for the NVESD VIS-LWIR spectrum, using WVU and CSU normalization, respectively. Overall, the assembled WVU dataset appears to be more challenging, perhaps because the images were not co-registered while being captured.

We gather that manifold learning is highly dependent on the data available for training and requires a good approximation of the underlying distribution of data. Data restraints are present, particularly in IR to visible FR where datasets are very limited in population. The collection and organization of such data, particularly data that has been co-registered, should be considered for the future. Also, the use of techniques such as neural networks that may be able to learn mappings by adjusting projection coefficients over the training set should be taken considered for future use as well.

Acknowledgements The authors would like to thank Dr. Mingsong Dou for his contributions in helping understand important concepts from the initial study [10]. The authors would also like to thank Dr. Shuowen Hu and the US Army Research Laboratory for granting us access to the NVESD dataset used in this work.

References

1. Ansari, A., Mahoor, M., Abdel-Mottaleb, M.: Normalized 3D to 2D model-based facial image synthesis for 2D model-based face recognition. In: IEEE GCC Conference and Exhibition (GCC), pp. 178–181 (2011)
2. Belhumeur, P., Hespanha, J., Kriegman, D.J.: Eigenfaces vs. fisherfaces: recognition using class specific linear projection. *IEEE Trans. Pattern Anal. Mach. Intell.* **19**(7), 711–720 (1997)
3. Bolme, D., Beveridge, J., Teixeira, M., Draper, B.: The CSU face identification evaluation system: its purpose, features and structure. In: Proceedings of International Conference on Vision Systems, pp. 301–311 (2003)
4. Bourlai, T., Kalka, N., Cao, D., Decann, B., Jafri, Z., Nicolo, F., Whitlam, C., Zuo, J., Adjero, D., Cukic, B., Dawson, J., Hornak, L., Ross, A., Schmid, N.A.: *Ascertaining Human Identity in Night Environments*. Princeton University Press, Princeton (2010a)
5. Bourlai, T., Kalka, N., Ross, A., Cukic, B., Hornak, L.: Cross-spectral face verification in the short wave infrared (SWIR) band (2010b)
6. Bourlai, T., Ross, A., Chen, C., Hornak, L.: A study on using middle-wave infrared images for face recognition. In: SPIE, Biometric Tech for Human Identification IX (2012)
7. Byrd, K.: Preview of the newly acquired NVESD-ARL multimodal face database. In: Proceedings of SPIE, vol. 8734, p. 34 (2013)
8. Dalal, N., Triggs, B.: Histograms of oriented gradients for human detection. In: Proceedings of IEEE Conference Computer Vision and Pattern Recognition, pp. 886–893 (2005)
9. Devijver, A.P., Kittler, J.: *Pattern Recognition: A Statistical Approach*. Prentice-Hall, London (1982)
10. Dou, M., Zhang, C., Hao, P., Li, J.: Converting thermal infrared face images into normal gray-level images. In: ACCV (2007)
11. Chang, H., Yeung, D., Xiong, Y.: Super-resolution through neighbor embedding. In: CVPR (2004)

12. Klare, B., Jain, A.: Heterogeneous face recognition: matching NIR to visible light images. In: ICPR, pp. 1513–1516 (2010)
13. Klare, B., Jain, A.: Heterogeneous face recognition using kernel prototype similarities. *IEEE Trans. Pattern Anal. Mach. Intell.* **35**(6), 1410–1422 (2013)
14. Klare, B., Li, Z., Jain, A.: Matching forensic sketches to mug shot photos. *IEEE Trans. Pattern Anal. Mach. Intell.* **33**(3), 639–646 (2011)
15. Lei, Z., Li, S.: Coupled spectral regression for matching heterogeneous faces. In: CVPR (2009)
16. Lei, Z., Liao, S., Jain, A., Li, S.: Coupled discriminant analysis for heterogeneous face recognition. *IEEE Trans. Inf. Forensics Secur.* **7**(6), 1707–1716 (2012)
17. Li, C., Su, G., Shang, Y., Li, Y., Xiang, Y.: Face recognition based on pose-variant image synthesis and multi-level multi-feature fusion. In: AMFG, pp. 261–275 (2007)
18. Lin, D., Tang, X.: Inter-modality face recognition. *Proc. Eur. Conf. Comput. Vis.* **3954**, 13–26 (2006)
19. Liu, Q., Tang, X., Jin, H., Lu, H., Ma, S.: A nonlinear approach for face sketch synthesis and recognition. In: CVPR, vol. 1, pp. 1005–1010 (2005)
20. Liu, W., Liu, J., Tang, X.: Bayesian tensor inference for sketch-based facial photo hallucination. In: IJCAI, pp. 2141–2146 (2007)
21. Lu, X., Hsu, R., Jain, A., Kamgar-Parsii, B.: Face recognition with 3D model-based synthesis. In: Proceedings of International Conference on Biometric Authentication (ICBA), pp. 139–146 (2004)
22. Majumdar, A., Ward, R.K.: Single image per person face recognition with images synthesized by non-linear approximation. In: International Conference on Image Processing, pp. 2740–2743 (2008)
23. Melzer, T., Reiter, M., Bischof, H.: Appearance models based on kernel canonical correlation analysis. *Pattern Recognit.* **36**, 1961–1971 (2003)
24. Mohideen, S.K., Perumal, S.A., Sathik, M.M.: Image de-noising using discrete wavelet transform. *Int. J. Comput. Sci. Netw. Secur.* **8**(1), 213–216 (2008)
25. Muller, K., Mika, S., Ratsch, G., Tsuda, K., Scholkopf, B.: An introduction to kernel-based learning algorithms. In: *IEEE Transactions on Neural Networks*, pp. 181–201 (2001)
26. Pietikinen, M.: Image analysis with local binary patterns. In: Proceedings of Scandinavian Conference on Image Analysis, pp. 115–118 (2005)
27. Roweis, S., Saul, L.: Nonlinear dimensionality reduction by locally linear embedding. *Science* **290**(5500), 2323–2326 (2000)
28. Sarfraz, M., Stiefelhagen, R.: Deep perceptual mapping for thermal to visible face recognition. In: British Machine Vision Conference (BMVC) (2016)
29. Selinger, A., Socolinsky, D.A.: Face recognition in the dark. In: CPRW, pp. 129–134 (2004)
30. Shuowen, H., Short, N., Gurram, P., Gurton, K., Reale, C.: FR Across the Imaging Spectrum, chap. 4 (2016)
31. Sirovich, L., Kirby, M.: Application of the Karhunen-Loeve procedure for the characterization of human faces. *IEEE Trans. Pattern Anal. Mach. Intell.* **12**(1), 103–108 (1990)
32. Socolinsky, D., Selinger, A., Neuheisel, J.: Face recognition with visible and thermal imagery. *CVIU* **91**, 72–114 (2003)
33. Tan, X., Triggs, B.: Enhanced local texture feature sets for face recognition under difficult lighting conditions. *Trans. Image Process.* **19**, 1635–1650 (2010)
34. Tang, X., Wang, X.: Face sketch recognition. *IEEE Trans. Circuits Syst. Video Technol.* **14**(1), 50–57 (2004)
35. Teixeira, M.: The bayesian intrapersonal/extrapersonal classifier. Ph.D. thesis, Colorado State University (2003)
36. Turk, M., Pentland, A.: Eigenfaces for recognition. *J. Cogn. Neurosci.* **3**(1), 71–86 (1991)
37. Viola, P., Jones, M.: Robust real-time face detection. *J. Comput. Vis.* **57**(2), 137–154 (2004)
38. Wang, R., Yang, J., Yi, D., Li, S.: An analysis-by-synthesis method for heterogeneous face biometrics. In: ICB (2009)
39. Wang, X., Tang, X.: Face photo-sketch synthesis and recognition. *IEEE Trans. Pattern Anal. Mach. Intell.* **31**(11), 1955–1967 (2009)
40. Wilder, J., Phillips, P., Jiang, C., Wiener, S.: Comparison of visible and infra-red imagery for face recognition. In: Automatic Face and Gesture Recognition, pp. 182–187 (1996)
41. Yi, D., Liu, R., Chu, R., Lei, Z., Li, S.: Partial face matching between near infrared and visual images in MBGC portal challenge. In: ICB (2007)
42. Yi, D., Liao, S., Lei, Z., Sang, J., Li, S.Z.: Partial face matching between near infrared and visual images in MBGC portal challenge. In: ICB, Springer, pp. 733–742 (2009)
43. Yoshitomi, Y., Miyaura, T., Tomita, S., Kimura, S.: Face identification using thermal image processing. In: WRHC, pp. 374–379 (1997)
44. Zhang, W., Wang, X., Tang, X.: Coupled information-theoretic encoding for face photo-sketch recognition. In: CVPR, pp. 513–520 (2011a)
45. Zhang, Z., Wang, Y., Zhang, Z.: Face synthesis from near-infrared to visual light via sparse representation. In: ICB (2011b)

Dr. Nnamdi Osia received the B.S. degree in computer engineering from the University of Maryland, Baltimore County (UMBC), in December 2009 and graduated with his Ph.D. in computer engineering in the Lane Department of Computer Science and Electrical Engineering, West Virginia University, Morgantown. He also received his certificate in Information Assurance and Biometrics from West Virginia University. Dr. Osia has a background in software and research engineering and development. Since 2013, his work has been published in conferences, journals and book chapters, and he has also given oral presentations at international workshops and conferences. His current research interests include biometrics, image processing, pattern recognition, and computer vision.

Dr. Thirimachos Bourlai is an *Assistant Professor* at LCSEE, WVU, an *Adjunct Assistant Professor* at the School of Medicine, Department of Ophthalmology, Eye Institute, WVU, and an *Adjunct Assistant Professor* at the Department of Forensic and Investigative Science since November 2014. In 2009, Dr. Bourlai founded and, since then, has been directing the Multi-Spectral Imagery Lab (MILab). He has several years of experience in the field of computer vision and has worked on, initiated, and directed several research projects. His research interests span various computer vision-related fields, with a current focus in the areas of image restoration of damaged documents and biometric systems using visible and hyper-spectra imaging sensors, useful for short- to long-range-based surveillance applications. His research projects have been funded by several government organizations including DoD, CITeR (NSF I/UCRC), FBI, etc.