On Matching Visible to Passive Infrared Face Images using Image Synthesis & Denoising

Nnamdi Osia Thirimachos Bourlai
West Virginia University
PO Box 6201, Morgantown, West Virginia 26506
nosia@mix.wvu.edu ThBourlai@mail.wvu.edu

***Abstract*— Performing a direct match between images from different spectra (i.e., passive infrared and visible) is challenging because each spectrum contains different information pertaining to the subject's face. In this work, we investigate the benefits and limitations of using synthesized visible face images from thermal ones and vice versa in cross-spectral face recognition systems. For this purpose, we propose utilizing canonical correlation analysis (CCA) and manifold learning dimensionality reduction (LLE). There are four primary contributions of this work. First, we formulate the cross-spectral heterogeneous face matching problem (visible to passive IR) using an image synthesis framework. Second, a new processed database composed of two datasets consistent of separate controlled frontal face subsets (VIS-MWIR and VIS-LWIR) is generated from the original, raw face datasets collected in three different bands (visible, MWIR and LWIR). This multi-band database is constructed using three different methods for preprocessing face images before feature extraction methods are applied. There are: (1) face detection, (2) CSU's geometric normalization, and (3) our recommended geometric normalization method. Third, a post-synthesis image denoising methodology is applied, which helps alleviate different noise patterns present in synthesized images and improve baseline FR accuracy (i.e. before image synthesis and denoising is applied) in practical heterogeneous FR scenarios. Finally, an extensive experimental study is performed to demonstrate the feasibility and benefits of cross-spectral matching when using our image synthesis and denoising approach. Our results are also compared to a baseline commercial matcher and various academic matchers provided by the CSU's Face Identification Evaluation System.**

I. Introduction

Face recognition (FR) has been an active and widely explored area of research over the last few decades, with a plethora of applications in military and law enforcement. However, a majority of FR research focuses primarily on visible band images (380-750 nm). Although visible band FR systems are considered to be efficient when face images are captured under controlled conditions, variation in pose, expression, and illumination is still considered to be a challenging problem. Unfortunately, FR based solely on visible band images may not be feasible in environmental conditions that are characterized by adverse lighting and conspicuous shadows (such as night-time environments [4], [28], [40]). Consequently, FR in the infrared (IR) spectrum has become an area of growing interest [31], [43], [41].

Differences in appearance arise between images sensed in the visible and active IR bands, primarily due to the properties of the object being imaged. The active IR spectrum consists of the Near IR band (0.7 - 0.9μm) and the lower range of the Short-Wave IR band (0.9 - 2.5μm). During data acquisition in the active IR band, a subject's face is usually actively illuminated using an external light source that can be detectable (i.e. in the case of the NIR band) or not (i.e. in the case of the SWIR band). The passive IR spectrum consists of the Mid-Wave IR (MWIR) (3 - 5μm), and Long-Wave IR (LWIR)] (7 - 14μm) bands. IR radiation in the form of heat is emitted from the target, in this particular case the subject's face, and detected by the camera sensor whenever data is acquired in the passive IR band. Passive IR sensors provide a significant capability of acquiring human biometric signatures under obscure environments without allowing the location of the sensor to be detected (as for example in the case of NIR sensors and the usage of active illumination). Combining the usage of passive IR sensors with other IR sensors (e.g. SWIR) can result in better performance of FR systems in environments that vary in illumination and standoff distances.

978-1-5090-4023-0/17/$31.00 ©2017 IEEE

A. Goals and Contributions

The contributions of this work are four-fold. First, we propose and formulate a visible to passive infrared face matching framework utilizing image synthesis. Second, two datasets of frontal face images consistent of paired VIS-MWIR and VIS-LWIR face images (using different methods for pre-processing prior to synthesis), are assembled. The datasets generated illustrate the challenges associated with our proposed cross-spectral face matching approach. One such challenge is the optimal placement of the synthesized dataset prior to matching (i.e. better used as the gallery or probe set?). Third, we propose a post-synthesis denoising methodology, which helps eliminate noise present in synthesized images and demonstrate face recognition accuracy is thus improved. Finally, by conducting an extensive experimental study we establish that images captured under the passive infrared spectrum can be matched to visible images, and vice-versa, with promising results; especially when our proposed pre-processing approach is employed before feature extraction and matching.An overview of the methodology proposed in this work is illustrated in Fig.1.

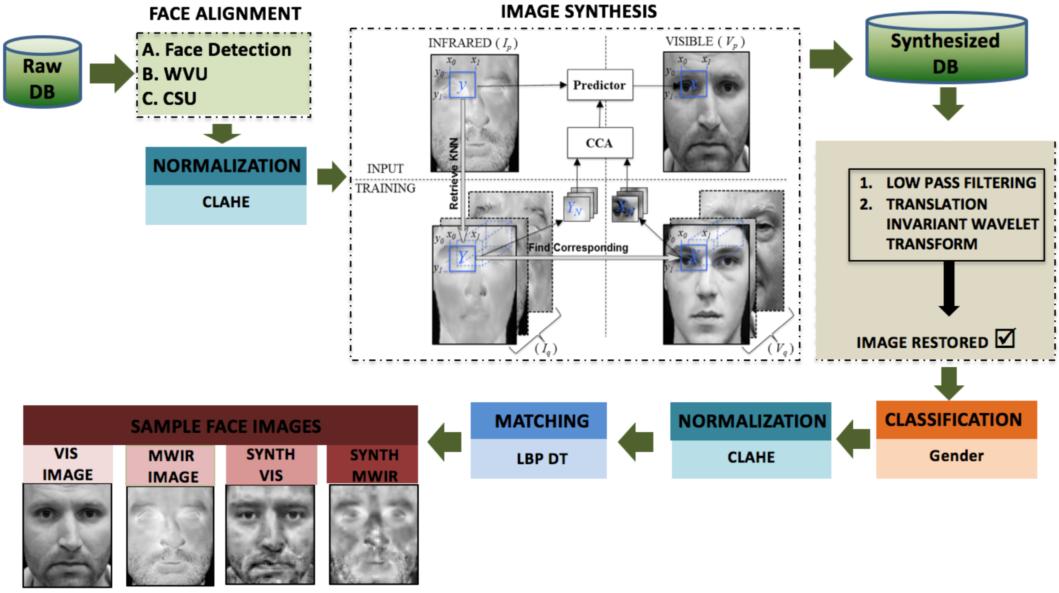

Fig. 1. Schematic of the proposed methodology for cross-spectral FR using image synthesis and denoising. The first step is face alignment using three different pre-processing methods. Next, image synthesis is carried out on each pre-processed database. Using our newly synthesized database, we perform image denoising and extract LBP DT features for FR.

B. Paper Organization

The rest of this paper is organized as follows. Sections 2, 3, 4, 5, and 6 describe related work, face-image synthesis, denoising and denoising, datasets and methodological steps, and experimental results. Conclusions and future work are described in Section 7.

II. Related Works

A. Heterogeneous FR

Tang et al. pioneered the work in this heterogeneous FR scenario with a number of approaches to synthesize a sketch using a visible image (or vice-versa) [33], [19], [37], [39]. A number of methods including, eigen-faces, LLE inspired local geometry preserving algorithms, kernel based nonlinear discriminant analysis (KNDA), Bayesian MAP framework, and a multiscale Markov Random Fields (MRF) model have been proposed in the literature to address various challenges of heterogenous FR matching scenarios. Aside from the generative transformation-based approaches, recent research in heterogeneous FR utilize approaches that are discriminative feature-based [44], [12], [14], [18], [16], [13], and have shown good accuracies for face matching in both the sketch-focused and NIR-based domains. Sarfraz et al. [27] 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. Kalka et al. [5] investigate the benefits and shortcomings of matching SWIR face images to visible images under controlled or uncontrolled conditions. study the problem of cross spectral face recognition in heterogeneous environments. Chen et al. [8] uses multiple sets of subspaces generated by patches sampled from visible and thermal face images and subjecting them to a series of transformations. Other implementations, on top of using non-linear dimensionality reduction and manifold learning, also use photometric normalization for optimal feature discrimination based on the spectrum of operation, and image reconstruction using the training data during the testing phase, in place of inferred features. An example of how image synthesis works is provided in Fig. 1. Please note that unlike other heterogeneous thermal-visible matching approaches, we use only the facial information (after face detection and normalization) for synthesis, denoising 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 [29].

B. Image Synthesis

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. Xu et al. present a cross-spectral dictionary learning approach using joint l_0 minimization in order to learn a mapping function between the VIS and NIR domain.

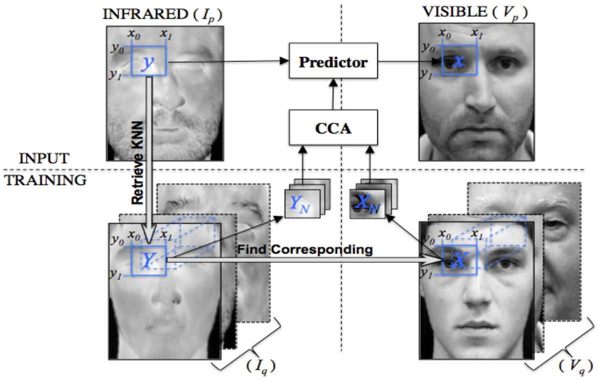

Fig. 2. Flow chart of image synthesis.

- **Subspace Methods:** In [21] 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. [42] 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 [20] a 3D generic face model is aligned with each frontal face image.

III. Face Image Synthesis

A. 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 *1*st pair of directions $\mathbf{w}_1$ and $\mathbf{v}_1$ that results in the greatest correlation between the projections x $= \mathbf{w_1}^{\mathrm{T}}\mathbf{x}$ and y $=$ $\mathbf{v_1}^{\mathrm{T}}\mathbf{y}, \max\rho(\mathbf{w_1}^{\mathrm{T}}\mathbf{x},\ \mathbf{v_1}^{\mathrm{T}}\mathbf{y}) \quad , s.t. \quad Var((\mathbf{w_1}^{\mathrm{T}}\mathbf{x} = 1)$ and $Var(\mathbf{v_1}^{\mathrm{T}}\mathbf{y} = 1)$, 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) $\mathbf{w}_k{}^{\mathrm{T}}\mathbf{x}$ and $\mathbf{v}_k{}^{\mathrm{T}}\mathbf{y}$ are not correlated to the previous *k-1* canonical variates; (2) the correlation between $\mathbf{w}_k{}^{\mathrm{T}}\mathbf{x}$ and $\mathbf{v}_k{}^{\mathrm{T}}\mathbf{y}$ is optimized under the constraints $Var((\mathbf{w_1}^{\mathrm{T}}\mathbf{x} = 1)$ and $Var(\mathbf{v_1}^{\mathrm{T}}\mathbf{y} = 1)$. Then $\mathbf{w}_k{}^{\mathrm{T}}\mathbf{x}$ and $\mathbf{v}_k{}^{\mathrm{T}}\mathbf{y}$ 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,

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

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

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 [22]. 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}.

B. 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 $\mathbf{W}^{(i)} = [\mathbf{w}_1, \mathbf{w}_2, \ldots, \mathbf{w}_k]$ and $\mathbf{V}^{(i)} = [\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_k]$ 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 preprocessed 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.

C. Reconstruction using Training Patches

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 $\mathbf{p} = \mathbf{Wf}$ 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,

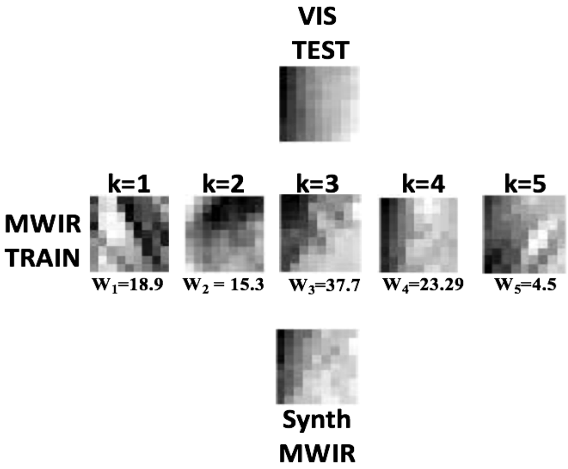

Fig. 3. Sample illustration of input VIS patch, and corresponding training MWIR patches, for k=5 nearest neighbor. The reconstructed and synthesized MWIR image using training patches and locally linear embedded weights.

$$\mathbf{p} = \mathbf{arg}_p\mathbf{min}||\mathbf{W}^T\mathbf{p} - \mathbf{f}||_2{}^2, \tag{3}$$

or to add an energy constraint,

$$\mathbf{p} = \mathbf{arg}_p\mathbf{min}||\mathbf{W}^T\mathbf{p} - \mathbf{f}||_2{}^2 + ||\mathbf{p}||_2{}^2. \tag{4}$$

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 **f** has to contain enough information about the original patch. The original patch can be recovered using LLE [26] 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 $\mathbf{p}_1, \mathbf{p}_2, \ldots, \mathbf{p}_k$ are the patches whose features $\mathbf{f}_1, \mathbf{f}_2, \ldots, \mathbf{f}_k$ are **f's** k nearest neighbors, and **f** is able to be recovered using neighboring features with **f = Fw**, where $\mathbf{F} = [\mathbf{f}_1, \mathbf{f}_2, \ldots, \mathbf{f}_k]$, $\mathbf{w} = [\mathrm{w}_1, \mathrm{w}_2, \ldots, \mathrm{w}_k]^\mathbf{T}$, we can reconstruct the original patch using **p = Pw,** where $\mathbf{P} = [\mathbf{p}_1, \mathbf{p}_2, \ldots, \mathbf{p}_k]$. Using a probe IR image, we partition it into small 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 **p = Pw** for reconstruction and then the patches will be combined into a VIS facial image. A sample illustration of the reconstruction process can be seen in Fig. 3 for K=5 nearest neighbors.

IV. Face Image Denoising

Unwanted noise is introduced into the image through the image synthesis process (see Fig. 2). Therefore, image denoising [23] is considered as a meaningful post-synthesis step that could help restore the structural and textural content of the image. Simple image filtering is not ideal for restoring useful image content because it can remove important frequency components in the pipeline. To help alleviate the challenge of effective removal of noise, linear denoising (e.g. filtering), and nonlinear denoising (e.g. thresholding) can be combined to account for both noise removal as well as restoration of the most important image features so that the matcher's accuracy is then expected to be improved.

V. Face Recognition

A. Datasets

The VIS-MWIR dual-band face dataset consists of 308 images (154 for probe and 154 for gallery) with four images in time per subject (77 subjects total). Visible images for both the VIS and MWIR datasets are extracted from videos captured in our laboratory, using a Canon EOS 5D Mark II camera. This digital SLR camera produces ultra-high resolution RGB color images or videos, with a resolution of 1920 × 1080 pixels. MWIR face images, which counterpart the visible dataset, are extracted from videos using a FLIR SC8000 MWIR camera. The infrared camera produces high definition thermal videos, with a resolution of 1024 × 1024.

The (2) VIS-LWIR dual-band (156 for probe and 156 for gallery) gains four images per subject (78 subjects total). The LWIR images are extracted using a FLIR SC600 LWIR camera. The science-grade infrared camera produces high-resolution LWIR images or videos, with a resolution of 640 × 480 pixels. The first 2 samples are utilized as gallery images, while the remaining 2 samples are the probe images. It is important to note that images between sensor pairs are not captured simultaneously and thus they are not co-registered (captured in both bands at the same time). Thus, our database is more challenging to work with when using our proposed patch-based synthesis and image denoising approach. We capture data by focusing the camera on the subject's complete head and shoulders. It is noteworthy that some of the subjects from both datasets do overlap, but contain different subjects so it difficult to say which spectrum of operation would be best for synthesis.

B. Methodological Steps

The salient stages of the proposed method 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. 4.
   - **Face Detection:** For the visible spectrum of our database, Viola & Jones face detection algorithm [36] is used to localize the spatial

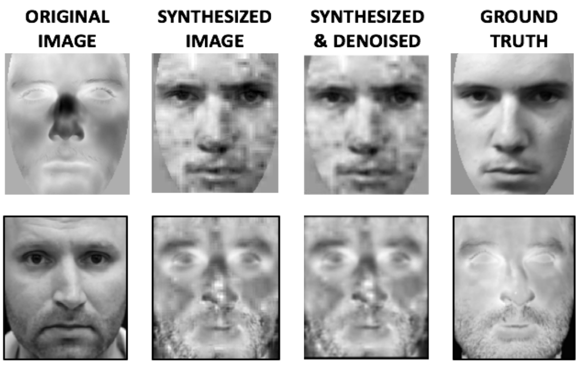

Fig. 4. Example original, synthesized, synthesized and denoised and ground truth images from two separate subjects. The subject on top row (MWIR to VIS) was normalized using CSU normalization, while the subject on the bottom row (VIS to MWIR) was normalized using our proposed normalization technique.

extent of the face and determine its boundary. 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 accurary 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:*** The methodology discussed in Section 3 is used. 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. By employing this algorithm, we process the datasets described in Section 5.1 and create their synthesized versions.

3) ***Image Denoising:*** The methodology detailed in [7] is used. We restore the synthesized images from the previous step using a combination of linear denoising and thresholding. Noniterative denoising methods (such as filtering and wavelet denoising with thresholding) allow for explicit numerical manipulation so that we are able to solve the noise problem in a single step. Ease of implementation and faster computation are the major advantages of noniterative methods.

4) ***Face Recognition Matcher:*** Both commercial and academic matchers, including the one provide by L1 systems and a set of matchers from the CSU face evaluation system, is utilized. While these matchers proved to be good, we also explored the usage of other matchers and distance metrics to determine which approach provides the best matching results after our proposed image synthesis and denoising. We found out that this is achieved if we utilize a variation of the *Local Binary Patterns* (LBP) method [32] for FR [5]. The LBP operator is an efficient, nonparametric, and unifying approach to traditional divergent models for analyzing texture that are statistical and structural based. A binary code is produced by thresholding the value of the center pixel with its value, for each pixel in an image [25].

VI. Empirical Evaluation

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 preprocessing 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 as the training data.

A. Baseline Experiments

We employ a set of baseline experiments (cross-spectral face matching) by using commercial and academic based software: 1) Commercial software Identity Tools (G8) provided by L1 Systems; 2) standard training-based face recognition methods provided by the CSU Face Identification Evaluation System [3], including *Principle Components*

TABLE I

Baseline Rank-1 FR results (%) for VIS-MWIR and VIS-LWIR Face Matching Experiments.

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

TABLE II

Proposed Approach Rank-1 FR results (%) for Synthesized VIS-MWIR and VIS-LWIR datasets using selected matcher (LBP DT).

Synthesized Rank-1 FR Accuracy (%) LBP DT				
Gallery	Probe	Face Detect	Proposed	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

Analysis (PCA) ([30],[35], [9]), *a combined Principle Components Analysis and Linear Discriminant Analysis algorithm (PCA+LDA)* [2], the *Bayesian Intrapersonal/Extrapersonal Classifier* (BIC) using either the Maximum likelihood (ML) or the Maximum *a posteriori* (MAP) hypothesis [34].

Using commercial matcher (G8), the rank-1 identification rate achieved is 40.26% for the VIS-MWIR dataset and 62.82% for the VIS-LWIR dataset. For the CSU academic matchers, the maximum rank-1 identification rate recorded is 19.23% for VIS-LWIR and 8.97% for VIS-MWIR using the LDA algorithm. These baseline results can be found in Table I.

B. Image Synthesis Experiments

We trained our synthesis algorithm 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 as the training data. 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. 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. Although not practical, our matching experiments after synthesis are tested using the synthesized images as both gallery and probe set. This was explored for each spectrum.

TABLE III

Rank-1 FR results (%) for Denoised Synthesized VIS-MWIR and VIS-LWIR datasets using selected matcher (LBP DT).

Proposed Approach Restored Synthesized Rank-1 FR Accuracy (%) LBP DT				
Gallery	Probe	Face Detect	Proposed	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**

We achieve a maximum rank-1 identification rate of 85.06% when using LBP (with the DT distance metric) matcher after synthesis for VIS-MWIR and a maximum rank-1 identification rate of 79.49% for the VIS-LWIR, using our proposed preprocessing approach. The rank-1 identification rates after image synthesis and denoising (pre-processed datasets) results can be found in Table II for each of our pre-processed IR datasets.

C. Image Denoising Experiments

In this experiment, we demonstrate that the combination of filtering and TI-denoising is essential for significantly improving 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 low-pass (LP) filtered and subsequently denoised. We optimize our proposed image denoising parameters, LP filter type and sigma value threshold used for TI-denosing, 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 window 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-denosing 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.

We achieve a maximum rank-1 identification rate of 81.17% when using LBP (with the DT distance metric) matcher after synthesis for VIS-MWIR and a rank-1 identification rate of 80.77% for VIS-LWIR, using our proposed and CSU normalization respectively. The results after denoising can be seen in Table III for our synthesized datasets. Identification rates (Rank-1 to Rank-5) can be seen for our collected data and proposed methodology (after denoising) when compared to classic academic matchers for practical scenarios (e.g. visible gallery) in Fig. **??**.

VII. Discussions & Conclusions

In this paper we study the problem of image synthesis as a means to bridge the informational gap between face images pertaining to two different spectral bands. Our study shows that image alignment is important in achieving higher FR

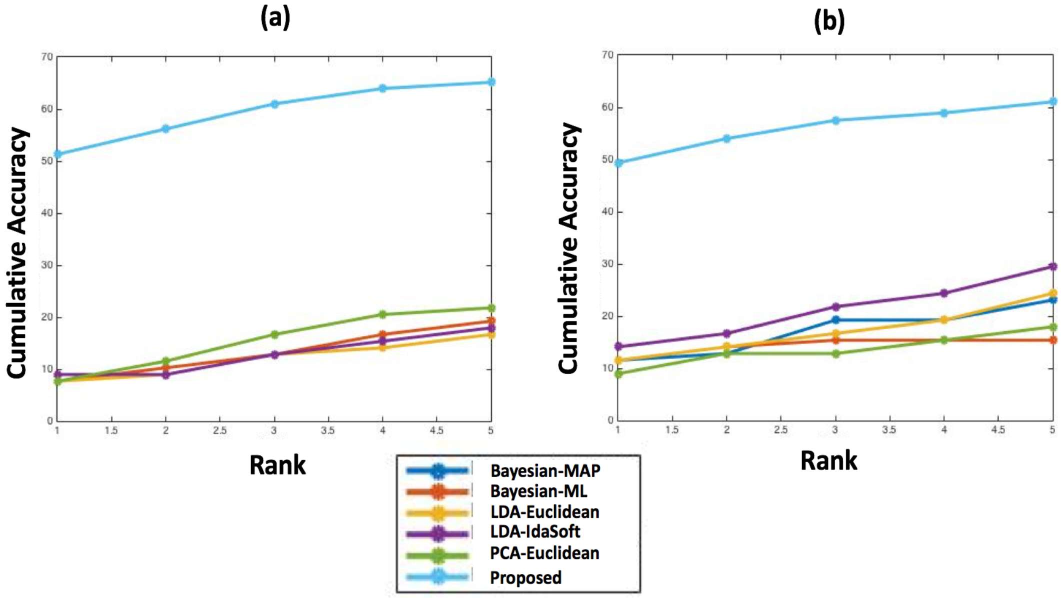

Fig. 5. (a) Identification rates (Rank-1 to Rank-5) for VIS gallery and MWIR probe. (b) Identification rates (Rank-1 to Rank-5) for VIS gallery and LWIR probe.

accuracy for the 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 VIS-MWIR and a maximum rank-1 identification rate of 79.49% for the VIS-LWIR spectrum, using our proposed geometric normalization. After denoising, we achieve a maximum rank-1 identification rate of 81.17% when using LBP DT matcher for VIS-MWIR and a rank-1 identification rate of 80.77% for VIS-LWIR, using our proposed and CSU normalization respectively. 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 of the amount of information contained 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 denoising 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 denoising 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 denoising step decreases face recognition accuracy for our proposed geometric normalization preprocessing step. Since we do not measure an exact model of noise, we are unable to come to any conclusions on noise in the image after denoising. The proposed approach is feasible given two large datasets that have a linear relationship, but is limited with data involving nonlinear correlation. In recent years it has been shown that using deep CNN-based features, and state-of-the-art hand-crafted features may already accommodate for cross-spectral variances during FR. When combining CCA that is based on deep CNNs, nonlinear correlation can be discovered, and distributions and functions can be learned for image synthesis. However, CNNs do have limitations as well, which includes high computational cost and need for a large amount of training data. Per a survey on HFR [24], CCA and LLE have also been used in some form or fashion for other synthesis approaches such as Sketch-Photo, VIS-NIR, and 2D-3D face. From our experiments, we gather that the proposed approach 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 into consideration as well.

Acknowledgments

This material is based upon work supported by the Center for Identification Technology Research and the National Science Foundation.

References

[1] A. Ansari, M. Mahoor, and M. Abdel-Mottaleb. Normalized 3d to 2d model-based facial image synthesis for 2d model-based face recognition. In *IEEE GCC Conf. and Exhibition (GCC)*, pages 178–181, 2011.

[2] P. Belhumeur, J. Hespanha, and D. J. Kriegman. Eigenfaces vs. fisherfaces: Recognition using class specic linear projection. *IEEE Trans. Patt. Anal. Mach. Intell.*, 19(7):711–720, July 1997.

[3] D. Bolme, J. Beveridge, M. Teixeira, and B. Draper. The csu face identification evaluation system: Its purpose, features and structure. In *Proc. International Conference on Vision Systems*, pages 301–311, 2003.

[4] T. Bourlai, N. Kalka, D. Cao, B. Decann, Z. Jafri, F. Nicolo, C. Whitelam, J. Zuo, D. Adjeroh, B. Cukic, J. Dawson, L. Hornak, A. Ross, and N. A. Schmid. *Ascertaining Human Identity in Night Environments*. Princeton Univeristy Press, 2010.

[5] T. Bourlai, N. Kalka, A. Ross, B. Cukic, and L. Hornak. *Cross-spectral Face Verification in the Short Wave Infrared (SWIR) Band*. 2010.

[6] T. Bourlai, A. Ross, C. Chen, and L. Hornak. A Study on using Middle-Wave Infrared Images for Face Recognition. *SPIE, Biometric Tech. for Human Identification IX*, April 2012.

[7] T. Bourlai, A. Ross, and A. Jain. Restoring degraded face images for matching faxed or scanned photos. *IEEE Transactions on Information Forensics and Security*, 6(2):371–384, June 2011.

[8] C. Chen and A. Ross. Matching thermal to visible face images using hidden factor analysis in a cascaded subspace learning framework. *Pattern Recognition Letters*, 72:25–32, March 2016.

[9] A. P. Devijver and J. Kittler. *Pattern Recognition: A Statistical Approach*. Prentice-Hall, 1982.

[10] M. Dou, C. Zhang, P. Hao, and J. Li. Converting thermal infrared face images into normal gray-level images. *ACCV*, 2007.

[11] H.Chang, D. Yeung, and Y. Xiong. Super-resolution through neighbor embedding. *CVPR*, 2004.

[12] B. Klare and A. Jain. Heterogeneous face recognition: Matching nir to visible light images. pages 1513–1516, 2010.

[13] B. Klare and A. Jain. Heterogeneous face recognition using kernel prototype similarities. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 35(6):1410–1422, 2013.

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

[15] Z. Lei and S. Li. Coupled spectral regression for matching heterogeneous faces. *CVPR*, 2009.

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

[17] C. Li, G. Su, Y. Shang, Y. Li, and Y. Xiang. Face recognition based on pose-variant image synthesis and multi-level multi-feature fusion. *AMFG*, pages 261–275, 2007.

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

[19] Q. Liu, X. Tang, H. Jin, H. Lu, and S. Ma. A nonlinear approach for face sketch synthesis and recognition. volume 1, pages 1005–1010, 2005.

[20] X. Lu, R. Hsu, A. Jain, B. Kamgar-Parsi, and B. Kamgar-Parsi. Face recognition with 3d model-based synthesis. In *Proc. Internat. Conf. on Biometric Authentication (ICBA)*, pages 139–146, July 2004.

[21] A. Majumdar and R. K. Ward. Single image per person face recognition with images synthesized by non-linear approximation. In *In: International Conference on Image Processing*, pages 2740–2743, 2008.

[22] T. Melzer, M. Reiter, and H. Bischof. Appearance Models Based on Kernel Canonical Correlation Analysis. *Pattern Recognition*, 36:1961–1971, 2003.

[23] S. K. Mohideen, S. A. Perumal, and M. M. Sathik. Image de-noising using discrete wavelet transform. *Int. J. Comput. Sci. Netw. Security*, 8(1):213–216, January 2008.

[24] S. Ouyang, T. Hospedales, Y. Song, X. Li, C. Loy, and X. Wang. A survey on heterogeneous face recognition: Sketch, infra-red, 3d and low-resolution. *Image and Vision Computing (IMAVIS)*, 56:28–48, December 2016.

[25] M. Pietikinen. Image analysis with local binary patterns. In *Proc. Scandinavian Conf. Image Anal.*, pages 115–118, June 2005.

[26] S. Roweis and L. Saul. Nonlinear dimensionality reduction by locally linear embedding. *Science*, 290(5500):2323–2326, 2000.

[27] M. Sarfraz and R. Stiefelhagen. Deep perceptual mapping for thermal to visible face recognition. In *British Machine Vision Conference (BMVC)*, 2016.

[28] A. Selinger and D. A. Socolinksy. Face recognition in the dark. *CPRW*, pages 129–134, June 2004.

[29] H. Shuowen, N. Short, P. Gurram, K. Gurton, and C. Reale. *FR Across the Imaging Spectrum*, chapter 4. Feb 2016.

[30] L. Sirovich and M. Kirby. Application of the karhunen-loeve procedure for the characterization of human faces. *IEEE Trans. Patt. Anal. Mach. Intell.*, 12(1):103–108, January 1990.

[31] D. Socolinsky, A. Selinger, and J.Neuheisel. Face recognition with visible and thermal imagery. *CVIU*, 91:72–114, 2003.

[32] X. Tan and B. Triggs. Enhanced local texture feature sets for face recognition under difficult lighting conditions. *Trans. Img. Proc.*, 19, June 2010.

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

[34] M. Teixeira. *The Bayesian Intrapersonal/Extrapersonal Classifier*. PhD thesis, Colorado State University, 2003.

[35] M. Turk and A. Pentland. Eigenfaces for recognition. 3(1):71–86, 1991.

[36] P. Viola and M. Jones. Robust real-time face detection. *Journal of Computer Vision*, 57(2):137–154, 2004.

[37] J. L. W. Liu, X. Tang. Bayesian tensor inference for sketch-based facial photo hallucination. In *IJCAI*, pages 2141–2146, 2007.

[38] R. Wang, J. Yang, D. Yi, and S. Li. An analysis-by-synthesis method for heterogeneous face biometrics. In *ICB*, 2009.

[39] X. Wang and X. Tang. Face photo-sketch synthesis and recognition. *IEEE Trans. Pattern Anal. Mach. Intell.*, 31(11):1955–1967, 2009.

[40] J. Wilder, P. Phillips, C. Jiang, and S. Wiener. Comparison of visible and infra-red imagery for face recognition. In *Automatic Face and Gesture Recognition*, pages 182–187, October 1996.

[41] D. Yi, S. Liao, Z. Lei, J. Sang, and S. Z. Li. Partial face matching between near infrared and visual images in mbgc portal challenge. In *ICB*, pages 733–742. Springer-Verlag, 2009.

[42] D. Yi, R. Liu, R. Chu, Z. Lei, and S. Li. Partial face matching between near infrared and visual images in mbgc portal challenge. In *ICB*, 2007.

[43] Y. Yoshitomi, T. Miyaura, S. Tomita, and S. Kimura. Face identification using thermal image processing. *WRHC*, pages 374–379, 1997.

[44] W. Zhang, X. Wang, and X. Tang. Coupled information-theoretic encoding for face photo-sketch recognition. pages 513–520, 2011.

[45] Z. Zhang, Y. Wang, and Z. Zhang. Face synthesis from near-infrared to visual light via sparse representation. *ICB*, 2011.