A Polarimetric Thermal Database for Face Recognition Research

Shuowen Hu[1] Nathaniel J. Short[2] Benjamin S. Riggan[1] Christopher Gordon[1]
Kristan P. Gurton[1] Matthew Thielke[1] Prudhvi Gurram[2] Alex L. Chan[1]

[1]U.S. Army Research Laboratory, 2800 Powder Mill Rd., Adelphi, MD 20783, USA
[2]Booz Allen Hamilton, 8283 Greensboro Dr., McLean, VA 20171, USA

Abstract

We present a polarimetric thermal face database, the first of its kind, for face recognition research. This database was acquired using a polarimetric longwave infrared imager, specifically a division-of-time spinning achromatic retarder system. A corresponding set of visible spectrum imagery was also collected, to facilitate cross-spectrum (also referred to as heterogeneous) face recognition research. The database consists of imagery acquired at three distances under two experimental conditions: neutral/baseline condition, and expressions condition. Annotations (spatial coordinates of key fiducial points) are provided for all images. Cross-spectrum face recognition performance on the database is benchmarked using three techniques: partial least squares, deep perceptual mapping, and coupled neural networks.

1. Introduction

Face recognition has been an active area of research for the past several decades, given its wide range of potential applications in the commercial, military, and government sectors. However, face recognition research and development have focused primarily on the visible spectrum. Recently, some researchers have evaluated and developed methods for face recognition in the infrared spectrum, particularly in the near-infrared (NIR, 0.74-1 μm wavelength), short-wave infrared (SWIR, 1-3 μm), and thermal infrared to a limited extent. The thermal infrared spectrum is composed of two bands: mid-wave infrared (MWIR, 3-5 μm) and longwave infrared (LWIR, 8-14 μm). While the phenomenology in the NIR and SWIR bands are reflection dominated, imaging in the MWIR and LWIR is typically emission dominated. Consequently, facial signatures acquired in the NIR and SWIR bands are more similar to the visible spectrum facial signature, while facial signatures collected in the thermal band appear significantly different from their visible spectrum counterparts. It is not until very recently that the polarization-state information of LWIR emissions has been exploited to enhance face recognition performance in the thermal band. Figure 1 shows the visible spectrum image of a subject and the corresponding conventional thermal and polarimetric thermal images.

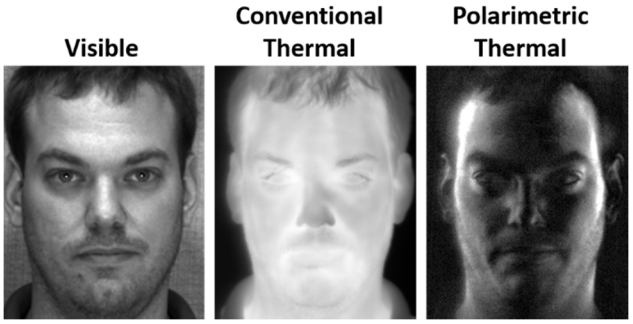

Figure 1: Visible spectrum, conventional thermal, and polarimetric thermal images of a subject.

A particular focus of face recognition research across the imaging spectrum has been on developing cross-spectrum, or heterogeneous, face recognition algorithms. The objective of cross-spectrum face recognition is to recognize the identity of an individual imaged in one spectral band (e.g. thermal probe) from a gallery database containing face imagery acquired in another band (e.g. visible spectrum database). For cross-spectrum face recognition, the gallery is usually taken to be in the visible spectrum, corresponding to the real-world scenario where government-maintained biometric databases and watch lists only contain visible spectrum face imagery. Given the similar reflective phenomenology in the visible, NIR, and SWIR bands, research on NIR-to-visible face recognition [1,2] and SWIR-to-visible face recognition [3,4] have achieved some measure of success. However, thermal-to-visible face recognition is significantly more challenging due to the difference in phenomenology between the thermal and visible spectra. Recently, several research groups [5-8] have developed techniques for and evaluated the performance of thermal-to-visible face recognition, achieving limited success. Although thermal-to-visible face recognition is highly challenging, it is also highly desirable as thermal imaging is relatively illumination invariant and is useful for providing a passive day and night face recognition capability.

Very recently, we have investigated and are continuing to explore polarimetric imaging in the LWIR for enhanced

cross-spectrum face recognition [9,10,13,14]. The polarization state information of thermal emissions contain geometric and textural facial details that are not present in conventional thermal face imagery [9], which can be used to improve face identification performance [10]. While many visible spectrum face databases exist, and there are a few face databases in the infrared bands, no polarimetric thermal face database had been available previously. In this work, we present a face database containing corresponding polarimetric thermal and visible spectrum imagery, and make it available to biometrics researchers to facilitate the development of cross-spectrum and multi-spectrum face recognition algorithms.

Section 2 of this paper provides a brief overview of polarimetric thermal imaging. Section 3 describes the sensor and experimental setup of the data collection, as well as the procedure used to extract the frames that compose the database. Section 4 details a baseline method and two state-of-the-art techniques that we used to benchmark cross-spectrum face recognition performance on this database. Section 5 presents the results, and Section 6 concludes the paper.

2. Overview of Polarimetric Thermal Imaging for Face Recognition

Both man-made and naturally occurring materials emit thermal infrared radiation that exhibit a preferential linear polarization state [9]. This preferential linear polarization is conjectured to originate by an anisotropy from the superposition of elemental thermal radiators near the surface-air interface of a material [11], which in turn causes the directional nature of the spectral emissivity [12]. By acquiring polarization-state imagery of the human face in the thermal spectrum, initial research has shown that additional textual and geometric information can be obtained, as compared to conventional intensity-based thermal face imagery [9].

The Stokes parameters S_0, S_1, S_2, and S_3 are commonly used to represent polarization-state information, defined in Eqs. 1-4, respectively. These Stokes parameters are determined on a pixel-by-pixel basis by measuring the radiant intensity transmitted through a polarizer and wave-plate pair that is rotated to various angles. We will refer to these Stokes parameters as Stokes images for the purposes of this work.

$$S_0 = I_0 + I_{90} \quad (\mathrm{W/sr\,cm^2}) \quad (1)$$

$$S_1 = I_0 - I_{90} \quad (\mathrm{W/sr\,cm^2}) \quad (2)$$

$$S_2 = I_{45} - I_{-45} \quad (\mathrm{W/sr\,cm^2}) \quad (3)$$

$$S_3 = I_R - I_L \quad (\mathrm{W/sr\,cm^2}) \quad (4)$$

In Eqs. 1-3, I_0, I_{90}, I_{45}, and I_{-45} represent the measured radiant intensity of the linear states at angles 0°, 90°, 45°, and -45°, respectively (relative to the horizontal). S_0 represents the conventional total intensity thermal image, while S_1 captures the horizontal and vertical polarimetric information, and S_2 captures the diagonal polarimetric information. Note that S_1 and S_2 capture orthogonal, yet complementary, polarimetric information. In Eq. 4, I_R and I_L are the right and left circularly polarized radiant states. In typical applications, S_3 is very small and usually taken to be zero [9]. The degree-of-linear-polarization (DoLP) describes the portion of an electromagnetic wave that is linearly polarized, as defined in Eq. 5.

$$\mathrm{DoLP} = \sqrt{S_1^2 + S_2^2} \Big/ S_0 \quad (5)$$

We first assessed polarimetric thermal face imagery from a visual perspective [9], and then proceeded to develop a feature extraction approach exploiting the Stokes imagery followed by support vector machine (SVM) classification to enhance cross-spectrum face recognition, demonstrating that polarimetric thermal-to-visible face identification significantly outperformed conventional thermal-to-visible face identification [10]. The polarization state information represented by S_1 and S_2 complements the conventional intensity only information represented by S_0, providing additional textural and geometric details that enhance recognition. Most recently, we developed neural network algorithms (based on deep perceptual mapping and coupled neural networks) to optimally combine polarimetric features followed by partial least squares (PLS) for classification, achieving the best results reported to date for polarimetric thermal-to-visible face recognition [13].

The Stokes images S_0, S_1, and S_2, can also be used to reconstruct a 3D representation of the facial surface. Yuffa et al. [14] developed a technique combining the Stokes images with Fresnel relations to estimate the surface normal at each pixel, followed by integration over the computed surface normals to generate a 3D facial image. Letting the surface normal be defined with respect to θ (angle between the surface normal and camera's imaging axis) and azimuth angle φ, the major challenge lies in resolving the inherent π ambiguity in the angle φ (i.e. whether the surface normal is pointing outwards or inwards is ambiguous). Yuffa et al. [14] introduced a set of constraints based on the boundary condition to partially overcome the π ambiguity in reconstructing the 3D facial surface. Being able to reconstruct the 3D facial surface from the Stokes images is potentially very useful for face recognition, enabling an off-frontal probe to be "frontalized" for pose-invariance when matching with gallery datasets that typically contain frontal or near-frontal face imagery.

3. Polarimetric Thermal Face Database

The face database presented here is the first of its kind, containing polarimetric LWIR imagery and simultaneously acquired visible spectrum imagery from a set of 60 distinct subjects. The following subsections describe the sensors (Section 3.1), the data collection protocol (Section 3.2), and provides details on the data preparation (Section 3.3).

Information on how to request this database is given in the Appendix.

3.1. Sensors

The polarimetric LWIR imager used to collect this database is shown in Figure 2, developed by Polaris Sensor Technologies. The imager is based on the division-of-time spinning achromatic retarder (SAR) design that uses a spinning phase-retarder mounted in series with a linear wire-grid polarizer [9]. This system, also referred to as a polarimeter, has a spectral response range of 7.5-11.1 μm, using a Stirling-cooled mercury telluride focal plane array with pixel array dimensions of 640×480. A Fourier modulation technique is applied to the pixel readout, followed by a series expansion and inversion to compute the Stokes images. Data were recorded at 60 frames per second (fps) for this database, using a wide FOV of 10.6°×7.9°. Prior to collecting data for each subject, a two-point non-uniformity correction (NUC) was performed using a Mikron blackbody at 20°C and 40°C, which covers the range of typical facial temperatures (30°C-35°C). Data was recorded on a laptop using custom vendor software.

Figure 2: Polarimetric LWIR imager used in this study.

An array of four Basler Scout series cameras was used to collect the corresponding visible spectrum imagery. Two of the cameras are monochrome (model # scA640-70gm), with pixel array dimensions of 659×494. The other two cameras are color (model # scA640-70gc), with pixel array dimensions of 658×494. Note that the pixel size for all four cameras are 7.4 μm × 7.4 μm. Objective lenses with different focal lengths were mounted on these cameras, to capture face images of different resolutions/interocular distances for the purposes of another study investigating the impact of resolution on face recognition performance. Table 1 lists the focal length and field of view for each Basler camera. Data from all four Basler cameras were streamed via GigE vision at 30 fps to a Quazar recording platform manufactured by Boulder Imaging, and recorded simultaneously.

Table 1. Basler Scout camera information

Camera	Mono or Color	Focal Length	Horizontal Field of View
VIS1	RGB	50 mm	5°
VIS2	Mono	16 mm	17°
VIS3	RGB	8.0 mm	34°
VIS4	Mono	4.5 mm	53°

3.2. Data collection protocol

The data collection protocol used to acquire the polarimetric thermal face database was reviewed and approved by the Institutional Review Board (IRB) of the U.S. Army Research Laboratory (ARL). Eighty subjects participated in this data collection, of which 60 subjects have consented to having their data be released for face recognition research. Data from the first 31 of the 60 consented subjects were acquired in the fall of 2014, while data from the remaining 29 subjects were acquired in the spring of 2016.

For the data collection, each subject was asked to sit in a chair and remove his or her glasses. A floor lamp with a compact fluorescent light bulb rated at 1550 lumens was placed 2 m in front of the chair to illuminate the scene for the visible cameras, as shown in Figure 3, and a uniform background was placed approximately 0.1 m behind the chair. Data was collected at three distances: Range 1 (2.5 m), Range 2 (5 m), and Range 3 (7.5 m). At each range, a baseline condition is first acquired where the subject is asked to maintain a neutral expression looking at the polarimetric thermal imager. A second condition, which we refer to as the "expressions" condition, was collected where the subject is asked to count out loud numerically from one upwards. Counting orally results in a continuous range of motions of the mouth, and to some extent, the eyes, which can be recorded to produce variations in the facial imagery. For each acquisition, 500 frames are recorded with the polarimeter (duration of 8.33 s at 60 fps), while 300 frames are recorded with each visible spectrum camera (duration of 10s at 30 fps).

Figure 3: Data collection setup showing the floor lamp and chair at Range 1 (2.5 m). In addition, a cardboard background was placed 0.1 m behind the chair.

3.3. Data preparation

The raw data recorded from the polarimeter are processed using custom software from Polaris Sensor Technologies to generate S_0, S_1, S_2, and DoLP sequences. For the S_1, S_2, and DoLP sequences from the baseline condition, 4 images are generated by averaging 24 consecutive frames at the 1 s, 3 s, 5 s, and 7 s marks, to improve the signal-to-noise ratio, as in [10]. No averaging is performed to extract the intensity-only S_0 images at these intervals. For the expressions condition where the subject counts orally, more images are extracted to sample the range of facial expressions. Specifically, 12 images are generated by averaging 24 consecutive frames at 12 uniform intervals for the S_1, S_2, and DoLP sequences from the expressions data. Again, no averaging is performed to extract the intensity-only S_0 images at these intervals. All averaged images are saved as 16-bit PNG files, and provided as part of the database release. Furthermore, the original 24-frame sequences are also provided in double precision (stored as Matlab .mat files) as part of the database release.

For the polarimetric thermal facial imagery, the inter-ocular distances are 87 pixels, 44 pixels, and 31 pixels at Range 1, Range 2, and Range 3, respectively. The database includes a set of visible facial imagery from the cameras that acquired similar interocular distances. Specifically, camera VIS1 data from Range 2 had the same 87 pixels interocular distance as the polarimetric thermal data from Range 1, camera VIS2 data from Range 1 had the same 44 pixels interocular distance as the polarimetric thermal data from Range 2, and camera VIS2 data from Range 2 had the same 31 pixels interocular distance as the polarimetric thermal data from Range 3. Four frames are extracted from these visible data sequences for the baseline condition, and twelve frames are extracted for the expressions condition, following the same procedure as for the polarimetric thermal data.

The set of polarimetric LWIR images was manually annotated. For the baseline condition, spatial coordinates of the centers of the eyes, tip of the nose, and center of the mouth were visually located and recorded. For the expressions condition, two additional coordinates are annotated: upper tip of the upper lip, and the lower tip of the lower lip. The visible spectrum imagery were automatically annotated, using PittPatt SDK v.5.2.2 developed by Pittsburgh Pattern Recognition, which produced accurate labeling for all visible spectrum imagery. Though multiple fiducial points are annotated/labeled, only the centers of the eyes are used for geometric normalization,

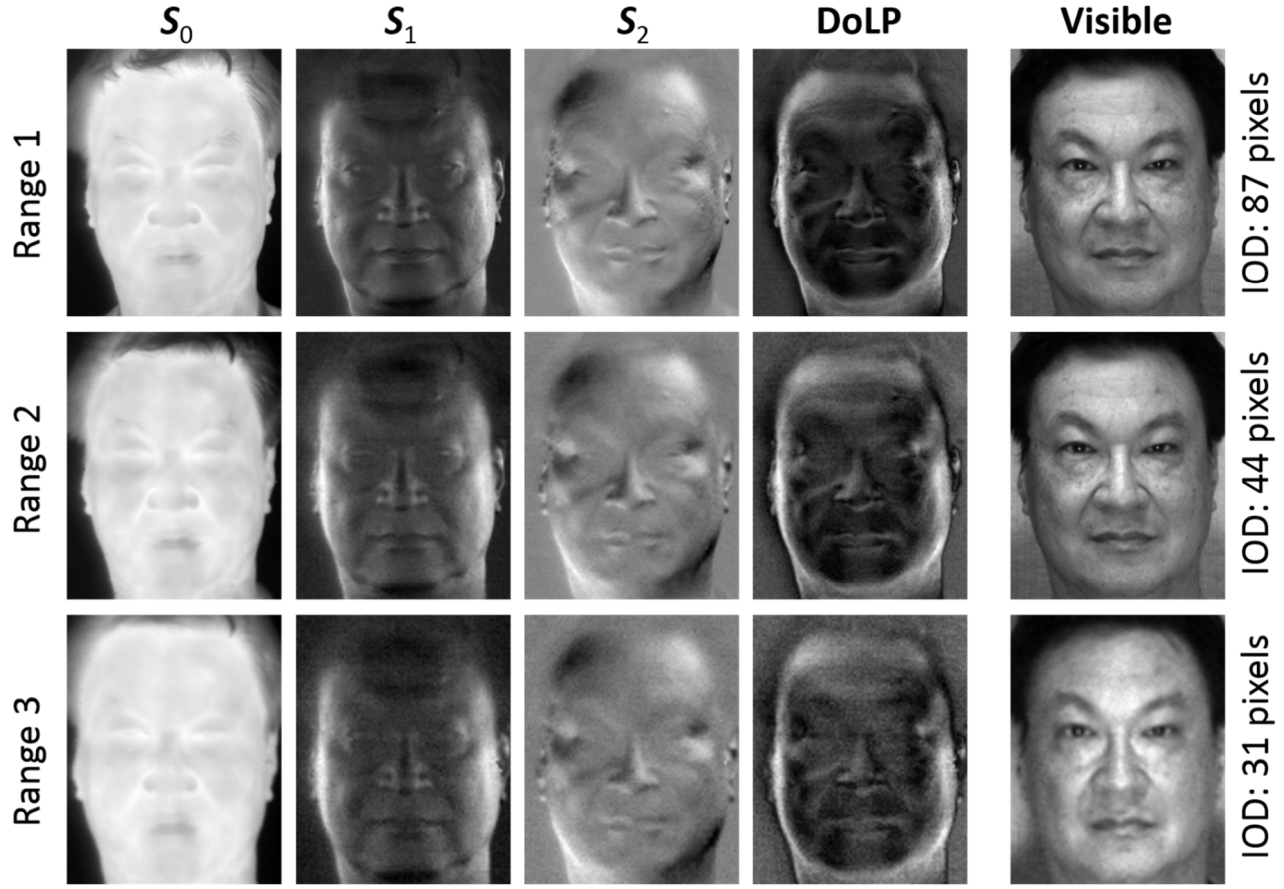

Figure 4. Example database imagery from the baseline condition, showing $\boldsymbol{S}_0$ (representing conventional thermal), $\boldsymbol{S}_1$, $\boldsymbol{S}_2$, and DoLP at three ranges (Range 1 – 2.5 m, Range 2 – 5.0 m, Range 3 – 7.5 m). Rightmost column shows example visible imagery from the database, selected from the cameras that produced similar interocular distance (IOD) as in the corresponding polarimetric imagery.

using the same procedure as in [7]. The procedure is as follows: (1) each face image is rotated so that the eye centers lie on a horizontal line, (2) the rotated image is then scaled so that the inter-ocular distance is 75 pixels, and (3) the scaled image is cropped to 250×200 pixels centered horizontally on the eyes (vertically placed at row 115). The last step produces a loose crop around the face, containing the facial outline along with the top of the head. The geometrically normalized imagery, both visible and polarimetric thermal, are provided as part of the database release.

4. Cross-Spectrum Face Recognition Evaluation

Three different techniques are used to quantify cross-spectrum face recognition performance on this database: partial least squares (PLS), deep perceptual mapping (DPM), and coupled neural networks (CpNN). The PLS approach [8] serves to provide a baseline measure of performance. The DPM and CpNN approaches have achieved the best results to date for polarimetric thermal-to-visible face recognition [13], and these two techniques are used here to quantify state-of-the-art performance on this database.

4.1. Deep perceptual mapping

The original DPM approach [15] consists of a multilayer neural network to perform regression on the image features directly, essentially finding a mapping that transforms visible face image features into thermal face image features. Specifically, Sarfraz and Stiefelhagen [15] used the dense scale invariant feature transform (SIFT), and trained a neural network to map visible SIFT features into the corresponding thermal SIFT feature representation (i.e. the predicted thermal SIFT features). Note that principal component analysis (PCA) was used to reduce the dimensionality of the feature vectors. After learning this mapping, Sarfraz and Stiefelhagen [15] used the following procedure for cross-spectrum face recognition. Given a gallery set of visible face images, visible SIFT features are extracted and then mapped to the corresponding thermal SIFT feature representation. Given an input thermal probe image for recognition, SIFT features were extracted from the thermal probe, and matching is performed by computing the cosine similarity between the thermal probe SIFT features and the predicted thermal SIFT features from the visible face images in the gallery [15].

For this work, we used an extended DPM model for polarimetric thermal imagery. As in [13], instead of the thermal SIFT features, we form polarimetric thermal SIFT features by concatenating SIFT features extracted from the S_0, S_1, and S_2 Stokes images, and learn the mapping from polarimetric thermal SIFT features to visible SIFT features. SIFT features were extracted from the visible, S_0, S_1, and S_2 Stokes imagery at two different scales by smoothing with Gaussian filters with standard deviations of 0.6 and 1.0. Patch size used was 20×20 pixels, with a stride of 8 pixels, resulting in a 128 dimensional SIFT feature vector for each patch. PCA was used to reduce the dimensionality to 64. The SIFT feature vector from each patch was augmented with its spatial location (i.e. row # and column #) to provide spatial dependence between the input and output, as in [13]. For a DMP model with k hidden layers and a single output layer (i.e. $n = k + 1$), the output of layer l is given by:

$$\boldsymbol{f}^{\ell}(\cdot) = \begin{cases} \sigma\left(\boldsymbol{W}^{\ell}\boldsymbol{f}^{\ell-1} + \boldsymbol{b}^{\ell}\right) & 1 \le \ell < n \\ \boldsymbol{W}^{\ell}\boldsymbol{f}^{\ell-1} & \ell = n \end{cases}, \quad (6)$$

where $\boldsymbol{W}$ is the weight matrix, $\boldsymbol{b}$ is the bias vector, and σ is the activation function. For this work, two hidden layers are used, each with 200 hyperbolic tangent units, similar to [15]. The weight matrices and bias vectors are learned by minimizing the following objective through mini-batch gradient descent:

$$J_{DPM}(\boldsymbol{W}^n, \ldots, \boldsymbol{W}^1, \boldsymbol{b}^n, \ldots, \boldsymbol{b}^1) = \sum_{i=1}^{M} \left\| \boldsymbol{v}_i - \widehat{\boldsymbol{v}}_i \right\|^2, \quad (7)$$

where $\boldsymbol{v_i}$ denotes the visible feature vector and $\widehat{\boldsymbol{v}}_i$ denotes the predicted visible feature vector from its polarimetric thermal counterpart. Cosine similarity is then used for matching.

4.2. Coupled neural network

In contrast to DPM, CpNN [13,16] performs an indirect regression, and extracts common latent features between the visible SIFT features and polarimetric thermal SIFT features by jointly learning two mappings. A CpNN therefore has two encoder networks, denoted by $\boldsymbol{F}_V(\boldsymbol{x}) = \boldsymbol{f}_V^k \circ \cdots \circ \boldsymbol{f}_V^2 \circ \boldsymbol{f}_V^1(\boldsymbol{x})$ and $\boldsymbol{F_T}(\boldsymbol{y}) = \boldsymbol{f}_T^k \circ \cdots \circ \boldsymbol{f}_T^2 \circ \boldsymbol{f}_T^1(\boldsymbol{y})$, where the subscript V represents the visible domain and the subscript T represents the polarimetric thermal domain, k is the number of hidden layers, and the output of each layer is given by $\boldsymbol{f}^{\ell} = \sigma(\boldsymbol{W}^{\ell} f^{\ell-1} + \boldsymbol{b}^{\ell})$. Note that unlike in DPM, dimensionality reduction is not necessary for the CpNN approach. A detailed description of the procedure to learn the parameters $\boldsymbol{W}_V^{\ell}, \boldsymbol{b}_V^{\ell}, \boldsymbol{W}_T^{\ell}, \boldsymbol{b}_T^{\ell}$ is given in [13]. For this work, each encoder has two hidden layers with 92 and 64 units. The trained encoders are used to extract the common latent descriptors from the visible SIFT features of the gallery images and the polarimetric thermal SIFT features of the probe set. The descriptors are normalized, and cosine similarity is used for matching.

4.3. Partial least squares

PLS was developed by Herman Wold and extended by Svante Wold, with its first applications in the social sciences [17]. Recently, PLS has been applied to face recognition in the visible spectrum [18] and for thermal-to-visible face recognition [8]. Let $\boldsymbol{X}_{m\times d}$ and $\boldsymbol{y}_{m\times 1}$ be the descriptor variable matrix and the corresponding label

vector, respectively, where m is the number of samples each with feature vector dimensionality of d. PLS regression finds a set of latent vectors $\boldsymbol{t}_i$ and $\boldsymbol{u}_i$, $i = 1 \ldots p$, that best predict $\boldsymbol{y}$:

$$\boldsymbol{X} = \boldsymbol{T}\boldsymbol{P}^T + \boldsymbol{X}_{res}, \quad (8)$$

$$\boldsymbol{y} = \boldsymbol{U}\boldsymbol{q}^T + \boldsymbol{y}_{res}. \quad (9)$$

In Eqs. 8 and 9, matrices $\boldsymbol{T}$ and $\boldsymbol{U}$ contain the extracted latent vectors, $\boldsymbol{P}$ and $\boldsymbol{q}$ are the loadings, and $\boldsymbol{X}_{res}$ and $\boldsymbol{y}_{res}$ are the residuals. The latent vectors can be computed by solving iteratively for a set of weight vectors $\boldsymbol{w}_i$:

$$\max \operatorname{cov}(\boldsymbol{t}_i, \boldsymbol{u}_i)^2 = \max_{\boldsymbol{w}_i} \operatorname{cov}(\boldsymbol{X}\boldsymbol{w}_i, \boldsymbol{y})^2 \ \forall i \ s.t. \ |\boldsymbol{w}_i| = 1, \quad (10)$$

Let $\boldsymbol{W}$ be the matrix of weight vectors $\boldsymbol{w}_i$, the PLS regression vector (i.e. PLS model) is computed as $\boldsymbol{\beta} = \boldsymbol{W}(\boldsymbol{P}^T\boldsymbol{W})^{-1}\boldsymbol{T}^T\boldsymbol{y}$. Given a visible gallery of N subjects, a PLS model is generated for each subject, using visible feature vectors from a given subject as positive samples, and visible feature vectors of the subjects from the training set as negative samples, similar to the procedure of [19]. The visible negative samples are augmented by a set of polarimetric thermal counter-examples (5 polarimetric thermal feature vectors from each subject in the training set). Let $\boldsymbol{f}$ be a polarimetric thermal probe feature vector, generated by averaging edge orientation features extracted the Stokes images per the procedure of [10]. A similarity score y_n^f can be computed for each subject's model (n=1,…,N) according to Eq. 11.

$$y_n^f = \bar{y}_n + \boldsymbol{\beta}_n^T \boldsymbol{f} \quad (11)$$

These similarity scores are used to compute face identification performance.

4.4. PLS∘DPM and PLS∘CpNN

Though DPM and CpNN are state-of-the-art approaches, face recognition performance can be further improved by using discriminative regression in the form of PLS for matching [13], instead of cosine similarity. As shown in Figure 5, the sequential processing chain consists of dense sampling (specifically dense SIFT), followed by DPM or CpNN, and then PLS (trained using the mapped features to improve the discrimination between match and non-match classes, as in [13]). These combined techniques are referred to as PLS∘DPM and PLS∘CpNN.

5. Experiments and Results

Of the 60 total subjects, data from 25 subjects were used for training and data from the remaining 35 subjects were used for testing. The training set was used to learn the DPM and CpNN mappings from the polarimetric thermal feature space to the visible spectrum feature space. For each modality (visible, thermal, and polarimetric thermal), there

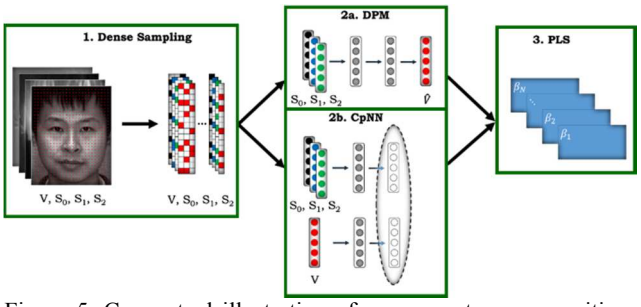

Figure 5. Conceptual illustration of cross-spectrum recognition framework consisting of SIFT feature extraction, followed by DPM or CpNN, and PLS for discriminative classification.

are four images per subject for the baseline condition and twelve images per subject for the expressions condition at each of the three ranges: Range 1 (2.5 m), Range 2 (5.0 m), and Range 3 (7.5m). The face resolutions in terms of inter-ocular distance at these ranges are 87, 44, and 31 pixels, respectively. Following geometric normalization as described in Section 3.3, the face regions were further cropped to 143×132 pixels, covering the core facial region that consists of the eyes, nose, and mouth. "Tight cropping" is used here, which is more practical from an operational standpoint compared to "loose cropping", as the tightly cropped facial region is less affected by changes in facial pose and hair style. In this work, as in [13], the Range 1 baseline visible images of the subjects in the testing set were used to form the gallery. This emulates the real-world scenario where biometric repositories typically contain higher resolution visible spectrum facial images. Features extracted from these Range 1 baseline visible images were used to build PLS models for classification/matching. Note that polarimetric thermal images from the training set were used as counter-examples to augment the negative samples from the testing set during model building to improve PLS performance [8]. To evaluate cross-spectrum face recognition performance, the probe set consists of four polarimetric thermal face images from each subject in the testing set for the baseline condition at each range, and twelve polarimetric thermal face images from each subject for the expressions condition at Range 1.

The performance evaluation was repeated 100 times, by randomly partitioning the subject pool into training and testing sets. For each trial, the same training set and testing set were used for PLS, DPM, and CpNN, and results are averaged over the 100 trials. Figure 7 shows the overall cumulative matching characteristics (CMC) curves for PLS, DPM, and CpNN, representing results aggregated over the three different ranges as well as the expressions data at Range 1. In additional to polarimetric thermal-to-visible face recognition performance (denoted with subscript polar in Figure 6's legend), we also show the results for conventional thermal-to-visible face recognition (denoted with subscript therm) for comparison. The

conventional thermal-to-visible results are generated using the same methodology as polarimetric thermal-to-visible, using S_0 to represent conventional thermal imagery and learning the corresponding DPM and CpNN mappings.

Three key findings can be summarized from the results shown in Figure 6. Firstly, learning the cross-spectrum feature mapping either directly using DPM or indirectly using CpNN significantly increases cross-spectrum face recognition performance compared to the baseline technique represented by PLS. Secondly, using discriminative classification in conjunction with cross-spectrum feature mapping (i.e. PLS∘DPM and PLS∘CpNN) outperforms DPM or CpNN alone (with the standard cosine similarity measure). Lastly, exploiting the polarization state information of thermal emissions enhances cross-spectrum face recognition performance compared to conventional "intensity-only" thermal-to-visible face recognition.

Table 2 tabulates the Rank-1 identification rates for a few different scenarios: overall (corresponding to Figure 6), Range 1 expressions, Range 1 baseline, Range 2 baseline, and Range 3 baseline. For PLS ∘ CpNN, exploiting polarization state information increased the Rank-1 identification rate by 2.58%, 6.52%, 6.01%, and 7.25%, compared to conventional thermal-to-visible face recognition, corresponding to Range 1 baseline, Range 1 expressions, Range 2 baseline, and Range 3 baseline, respectively. For PLS∘DPM, exploiting polarization state information increased the Rank-1 identification rate by 2.90%, 6.67%, 5.27%, and 6.77%, compared to conventional thermal-to-visible face recognition, corresponding to Range 1 baseline, Range 1 expressions, Range 2 baseline, and Range 3 baseline, respectively. This suggests that polarization state information is more beneficial for cross-spectrum face recognition under more challenging scenarios (e.g. expressions, longer range),

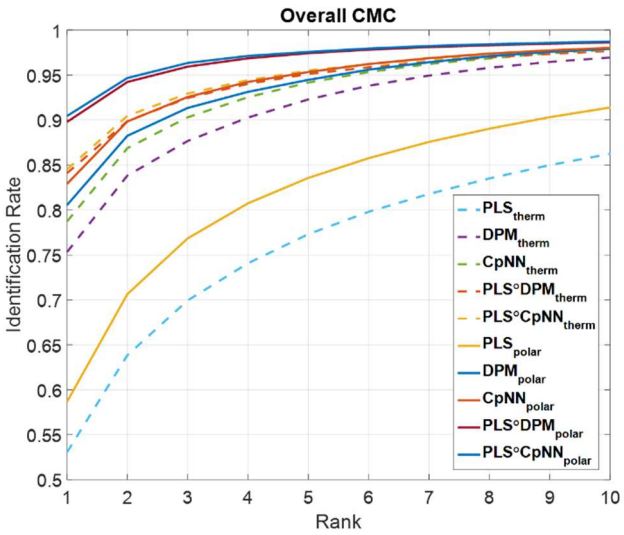

Figure 6: Overall cumulative matching characteristics curves from testing PLS, DPM, and CpNN using polarimetric (polar) and thermal (therm) probe samples, matching against a visible spectrum gallery.

which is unsurprising, as the conventional thermal-to-visible Rank-1 identification rate was already very high under the Range 1 baseline scenario (94.17% for PLS∘DPM and 93.88% for PLS ∘ CpNN). Another interesting observation from Table 2 is that while PLS∘CpNN has very similar performance to PLS∘DPM in almost all scenarios (within 1%, except for only Range 3), CpNN typically outperformed DPM by 1%-3%. This suggests that the indirect mapping to a common subspace is more effective than the direct mapping. However, the use of a discriminative classifier (e.g. PLS) in conjunction with DPM helps recover this deficit, resulting in similar performance between PLS∘DPM and PLS∘CpNN.

Table 2. Rank-1 identification rate of evaluated algorithms for cross-spectrum face recognition using polarimetric thermal and thermal probe imagery.

Scenario	*Rank-1 Identification Rate*					
	Probe	**PLS**	**DPM**	**CpNN**	**PLS∘DPM**	**PLS∘CpNN**
Overall	*Polar*	0.5867	0.8054	0.8290	0.8979	0.9045
	Therm	0.5305	0.7531	0.7872	0.8409	0.8452
Expressions	*Polar*	0.5658	0.8324	0.8597	0.9565	0.9559
	Therm	0.6276	0.7887	0.8213	0.8898	0.8907
Range 1 Baseline	*Polar*	0.7410	0.9092	0.9207	0.9646	0.9646
	Therm	0.6211	0.8778	0.9102	0.9417	0.9388
Range 2 Baseline	*Polar*	0.5570	0.8229	0.8489	0.9105	0.9187
	Therm	0.5197	0.7532	0.7904	0.8578	0.8586
Range 3 Baseline	*Polar*	0.3396	0.6033	0.6253	0.6445	0.6739
	Therm	0.3448	0.5219	0.5588	0.5768	0.6014

6. Conclusion

We present in this paper a polarimetric thermal face database, the first of its kind, for face recognition research. This database was acquired using a polarimetric longwave infrared imager, specifically a division-of-time spinning achromatic retarder system, which acquires geometric and textural facial details not available in conventional thermal face imagery. Three techniques are used to benchmark cross-spectrum face recognition performance: PLS, DPM, and CpNN. PLS results serve as the baseline performance measure, while DPM and CpNN are state-of-the-art approaches that, when used in conjunction with a discriminative classifier such as PLS, provide the current "gold standard" performance on this polarimetric thermal database. The motivation behind the release of this database is to help promote research on polarimetric and multi-spectrum face recognition. As no previous polarimetric thermal face database existed, our intention is to first provide a "pristine" database collected under controlled conditions. We recognize the need for additional data collected under more unconstrained real-world settings, and intend to do so in the near future, as sensor and algorithm development progresses.

Appendix

Requests for the database can be made by contacting the following points of contact (POCs) at the U.S. Army Research Laboratory:

- Matthew Thielke
  matthew.d.thielke.civ@mail.mil
- Shuowen (Sean) Hu
  shuowen.hu.civ@mail.mil

Each request will be vetted by the POCs to be for valid scientific research, and requestors will be asked to sign a database release agreement. The standard dataset described in Section 3.3 will then be made available to the requestor. Custom data requests may be accommodated on an individual basis.

Acknowledgements

The authors would like to thank Greg Samples and Patrick Rauss for their help in collecting the database, and Mark Acosta for his help in processing the data. Special thanks to David Chenault and Larry Pezzaniti at Polaris Sensor Technologies for sensor-related discussions.

References

[1] D. Yi, R. Liu, R.F. Chu, Z. Lei, and S.Z. Li, "Face matching between near infrared and visible light images," in Advances in Biometrics: Lecture Notes in Computer Science, vol. 4642, pp. 523-530, 2007.

[2] B. Klare and A.K. Jain, "Heterogeneous face recognition: matching NIR to visible light images," in Proc. Int. Conf. on Pattern Recognition, pp. 1513-1516, 2010.

[3] T. Bourlai, N. Kalka, A. Ross, B. Cukic, and L. Hornak, "Cross-spectral face verification in the short wave infrared (SWIR) band," in Proc. Int. Conf. on Pattern Recognition, pp. 1343-1347, 2010.

[4] F. Nicolo and N.A. Schmid, "Long range cross-spectral face recognition: matching SWIR against visible light images," IEEE Trans. on Information Forensics and Security, vol. 7, no. 6, 2012.

[5] J. Choi, S. Hu, S.S. Young, and L.S. Davis, "Thermal to visible face recognition," in Proc. SPIE DSS, vol. 8371, 2012.

[6] T. Bourlai, A. Ross, C. Chen, and L. Hornak, "A study on using mid-wave infrared images for face recognition," in Proc. SPIE DSS, vol. 8371, 2012.

[7] B.F. Klare and A.K. Jain, "Heterogeneous face recognition using kernel prototype similarity," IEEE Trans. on Pattern Analysis and Machine Intelligence, vol. 35, no. 6, pp. 1410-1422, 2013.

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

[9] Gurton K, Yuffa A, Videen G, "Enhanced facial recognition for thermal imagery using polarimetric imaging", Optics. Letters, vol. 39, no. 13, pp. 3857-3859, 2014.

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

[11] P.P. Feofilov, *The Physical Basis of Polarized Emission*, Consultants Bureau, 1961.

[12] R. Siegel and J.R. Howell, *Thermal Radiation Heat Transfer*, McGraw-Hill, 1981.

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

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

[15] M. S. Sarfraz, R. Stiefelhagen, "Deep Perceptual Mapping for Thermal to Visible Face Recognition", in Proc. British Machine Vision Conference, Sept. 2015.

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

[17] I. Helland, "Partial least squares regression," in Encyclopedia of Statistical Sciences (Wiley, 2006), pp. 5957–5962.

[18] W. R. Schwartz, H. Guo, J. Choi, and L. S. Davis, "Face identification using large feature sets," IEEE Trans. Image Processing, vol. 21, 2245–2255, 2012.