Improving cross-modal face recognition using polarimetric imaging

Nathaniel Short,[1,2,*] Shuowen Hu,[1] Prudhvi Gurram,[1,3] Kristan Gurton,[1] and Alex Chan[1]

[1]*U.S. Army Research Laboratory, 2800 Powder Mill Rd., Adelphi, Maryland 20783, USA*

[2]*Booz Allen Hamilton, 8283 Greensboro Dr, McLean, Virginia 22102, USA*

[3]*MBO Partners, 13454 Sunrise Valley Drive, Suite 300, Herndon, Virginia 20171, USA*

**Corresponding author: nathaniel.j.short2.ctr@mail.mil*

Received November 18, 2014; revised January 2, 2015; accepted January 3, 2015; posted January 22, 2015 (Doc. ID 226861); published March 3, 2015

We investigate the performance of polarimetric imaging in the long-wave infrared (LWIR) spectrum for cross-modal face recognition. For this work, polarimetric imagery is generated as stacks of three components: the conventional thermal intensity image (referred to as S_0), and the two Stokes images, S_1 and S_2, which contain combinations of different polarizations. The proposed face recognition algorithm extracts and combines local gradient magnitude and orientation information from S_0, S_1, and S_2 to generate a robust feature set that is well-suited for cross-modal face recognition. Initial results show that polarimetric LWIR-to-visible face recognition achieves an 18% increase in Rank-1 identification rate compared to conventional LWIR-to-visible face recognition. We conclude that a substantial improvement in automatic face recognition performance can be achieved by exploiting the polarization-state of radiance, as compared to using conventional thermal imagery.

OCIS codes: (110.0110) Imaging systems; (100.5010) Pattern recognition; (040.3060) Infrared; (110.5405) Polarimetric imaging.

http://dx.doi.org/10.1364/OL.40.000882

Automatic face recognition has a wide range of applications in the commercial, military, and government sectors, spanning from people tagging in social networking websites to surveillance for homeland security. Face recognition research has predominantly focused on the visible spectrum, addressing challenges such as illumination variations, pose, and image resolution. However, for surveillance during nighttime, the lack, or absence, of illumination prevents cameras operating in the visible-light spectrum from being used discreetly and effectively. Thermal imaging measures radiation in the mid-wave infrared (MWIR) and long-wave infrared (LWIR) spectra, which is naturally emitted by living tissue, and therefore is a highly practical imaging modality for nighttime operation. However, as most databases and watch lists only contain facial imagery in the visible spectrum, the challenge is to match an unknown thermal probe image of an individual's face to a set of known visible gallery images. This is referred to as cross-modal or heterogeneous face recognition, seeking to match probe face images acquired in one imaging modality to gallery face images from a different imaging modality.

Several recent efforts have attempted to address cross-modal, thermal-to-visible face recognition [1–3]. Due to the large modality gap caused by differences in phenomenology (reflectance for visible imaging and emittance for thermal imaging), the measured visible face signatures are very different from the thermal face signatures. The proposed techniques in [1–3], consisting of preprocessing, feature extraction, and classification, have been met with limited success. Bourlai *et al.* [3] achieved the highest level of accuracy; however, identification performance was less than 55% for thermal-to-visible face recognition. Thermal-to-visible face recognition algorithm performance may be fundamentally limited by the degree of correlation between the visible and thermal facial signatures, due to phenomenology and the lower spatial resolution in the thermal spectrum arising from the longer wavelength.

A growing area of interest among researchers is polarimetric imaging in the thermal spectrum, which is sensitive to changes in surface texture and geometry. The work by Gurton *et al.* [4], to the best of the authors' knowledge, is the first study in literature presenting facial imagery acquired in polarimetric LWIR, but it did not examine the contribution of polarization-state information to face recognition performance. This paper is a follow-up of Ref. [4], characterizing the facial information present in polarimetric LWIR and examining its performance for cross-modal face recognition. We demonstrate that a substantial improvement in automatic face recognition performance can be achieved by exploiting polarimetric information. Figure 1 shows facial signatures of a single subject acquired using visible, conventional LWIR, and polarimetric LWIR sensors, illustrating the signature differences.

The spectral emissivity in the thermal IR spectrum is directional in nature due to preferential linear polarization of the radiation emitted by both naturally occurring and man-made materials in this spectrum [5]. Hence, the polarimetric images generated based on each polarization

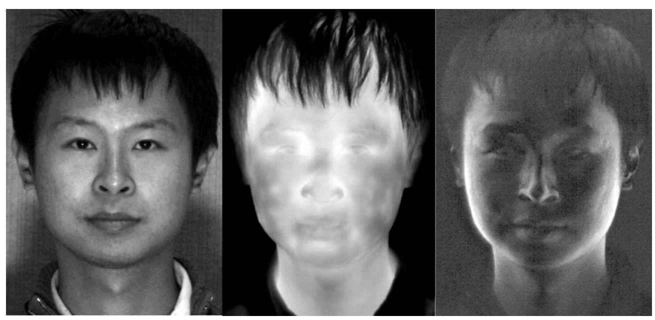


Fig. 1. Visible (left), conventional thermal (middle), and polarimetric-based (right) images of the same subject.

state contain texture and geometry details that were previously unavailable in the conventional thermal imagery. The polarization states for emitted or reflected light are described using the Stokes parameters S_0, S_1, and S_2, from which the Degree of Linear Polarization (DoLP) can be calculated [6]. These parameters are found by measuring radiant intensities of the linear states of polarization at angles 0°, 90°, +45°, and −45°, relative to the vertical axis of the focal plane. Each of these Stokes parameters is calculated on a pixel-by-pixel basis to construct 2D Stokes images. While the S_0 image represents the conventional thermal LWIR image without any polarization, S_1 and S_2 represent complementary, orthogonal polarimetric information, which can provide facial details that facilitate cross-modal face recognition.

The polarimetric LWIR face imagery used in this work was collected by a division-of-time spinning achromatic retarder (SAR) system from Polaris Sensor Technologies [4]. The system consisted of a Stirling-cooled mercury cadmium telluride FPA, with pixel array dimensions of 640×480, and a spectral response range of 7.5–11.1 μm. A sequence of 32-bit images is recorded at 60 Hz frame rate. A Fourier modulation technique is applied to the pixel readout, and a series expansion is calculated and inverted, which yields the coefficients used to generate the Stokes and DoLP images.

Since the visible and polarimetric face signatures exist in different spectra/domains, preprocessing steps must be applied to accentuate distinguishable and correlated facial features between the two domains. Here, we focus on the edges present in regions around the key facial features, which include the eyes, nose, and mouth. To accentuate these details and to reduce high- and low-frequency noise, a band-pass filter in the form of a difference of Gaussians (DoG) filter is applied to the visible and Stokes images. DoG preprocessing is defined in Eq. (1), and involves the convolution of an image with the difference of two Gaussian kernels with different bandwidth parameters:

$$D(x, y, \sigma_0, \sigma_1) = [G(x, y, \sigma_1) - G(x, y, \sigma_2)] * I(x, y), \quad (1)$$

where D is the DoG filtered image, * is the convolution operator, G is the Gaussian kernel, defined in Eq. (2), and σ_1 and σ_2 control the bandwidth of the filter.

$$G(x, y, \sigma) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{x^2+y^2}{2\sigma^2}} \quad (2)$$

Features are extracted from each of the DoG filtered images using the histogram of oriented gradients (HOG) feature representation [7]. The HOG feature is used because it provides robust encoding of distinguishable edge magnitude as well as orientation information for both visible and polarimetric face signatures. Dimensionality reduction is then performed using principal component analysis (PCA), and the resulting features are used for classification and recognition of faces. Here, we use the well-known support vector machine (SVM) [8] for classification. SVMs are binary classifiers, which discriminate between two classes by building a hyperplane that maximizes the margin between the two classes. To handle highly non-linear multi-modal data distributions, a "kernel trick" is used to map the data to a higher-dimensional Reproducing Kernel Hilbert space (RKHS), and a classifier is built in the RKHS [9]. In this work, a Gaussian radial basis function (RBF) kernel is used. Since the face recognition problem is a multi-class problem, we use "one-vs.-rest" strategy [9], where the samples belonging to a single subject form one class and the samples belonging to the rest of the subjects are considered to be the second class. Therefore, for N subjects, we train N "one-vs.-rest" SVM classifiers. Since this work focuses on cross-modal face recognition (i.e., recognizing a polarimetric probe/test image from a visible gallery/training dataset), each SVM classifier is trained only on the visible image set. Given a polarimetric probe image, N scores are obtained using each of the SVM classifiers. The scores are compared, and the gallery subject corresponding to the classifier returning the maximum score is identified as the top match.

A dataset of 20 subjects was collected for experimentation. Each subject was asked to remain still for eight seconds while visible and polarimetric thermal imagery was acquired at a distance of 2.5 meters. Software provided by Polaris Sensor Technologies was used to generate the Stokes images from the polarimetric data. Since the conventional LWIR radiometric data is represented by the S_0 image, baseline comparison is conducted using the S_0 image. Four samples of each subject were extracted from the visible and polarimetric LWIR video sequences. To improve the signal-to-noise ratio of the polarimetric data for S_1, S_2, and DoLP, each sample is an average of a 24-frame sequence.

Prior to training and testing of the cross-modal face recognition algorithm, face images must be transformed (i.e., aligned) to a common spatial coordinate system, referred to as canonical coordinates. First, corresponding fiducial points (corner of eyes, tip of nose, and center of mouth) are selected in each visible and polarimetric face image. Then, a spatial transformation is computed using the defined fiducial points to align the visible and polarimetric images to the canonical coordinates, where the fiducial points are in fixed positions. Figure 2 shows an example of a single subject from the database used in this work. The top row of Fig. 2 shows grayscale intensity-based images after aligning to canonical coordinates and cropping. The middle row shows faces after DoG preprocessing. The bottom row provides a visual representation of the extracted HOG features, where vectors representing directions 0 through 180 degrees are shown for each local region. The grayscale intensity of the vector indicates the magnitude of the local gradient in the direction of the vector.

As can be observed in Fig. 2, the Stokes images provide complementary information about the key structures of the human face. For example, the S_0 image provides highly correlated details around the ocular region (eyes and eyebrows) between the thermal and visible faces. In contrast, S_1 and S_2 provide a higher degree of correlation to the visible face signature around the nose and mouth regions. We therefore propose a composite representation derived from the feature-level fusion of the HOG representations of each Stokes image. The fusion approach combines local edge magnitude and directional

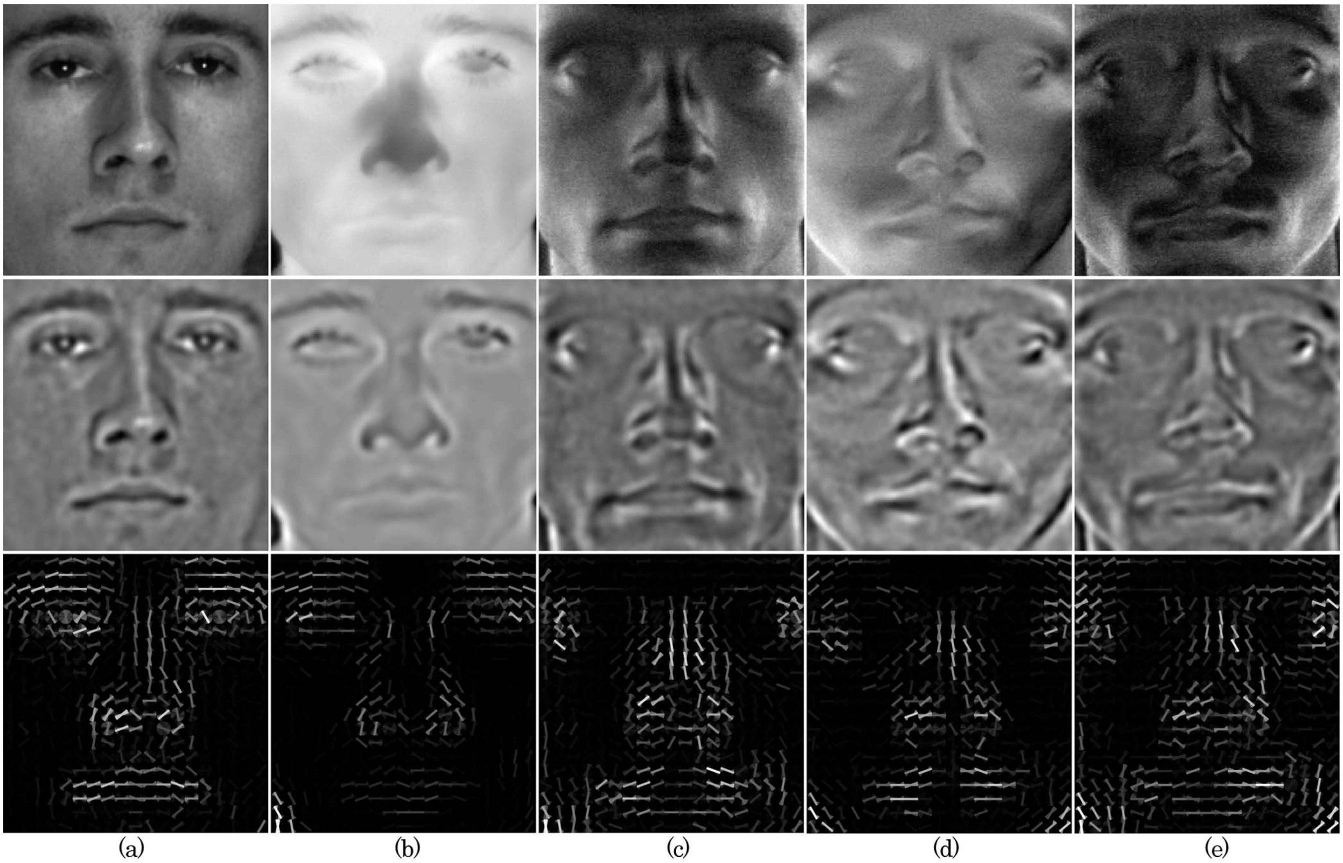


Fig. 2. Single subject as viewed in visible (a) and thermal (b-e) spectra. (Top) The Stokes images, S_0, S_1, S_2, are shown in (b), (c), and (d), respectively. The DoLP image is shown in (e). (Middle) Images after DoG filtering, where high/low spatial frequencies have been attenuated. (Bottom) HOG feature representations, where grayscale intensity represents magnitude of local edge direction.

information by averaging the values across each spatial region of the HOG feature sets corresponding to the three Stokes images. Figure 3 provides a visual illustration of the proposed composite feature representation. It can be seen that local regions across all the Stokes images have been combined into the composite image, providing a higher

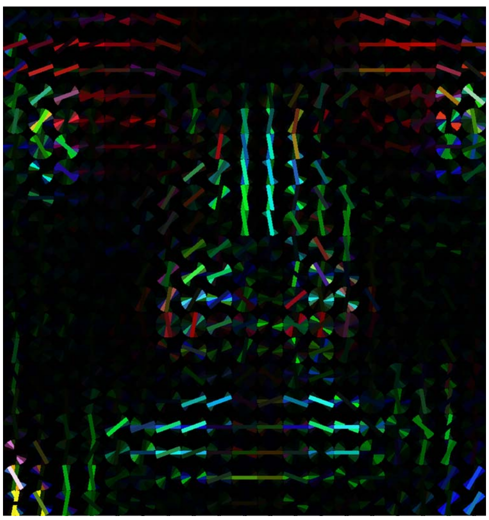
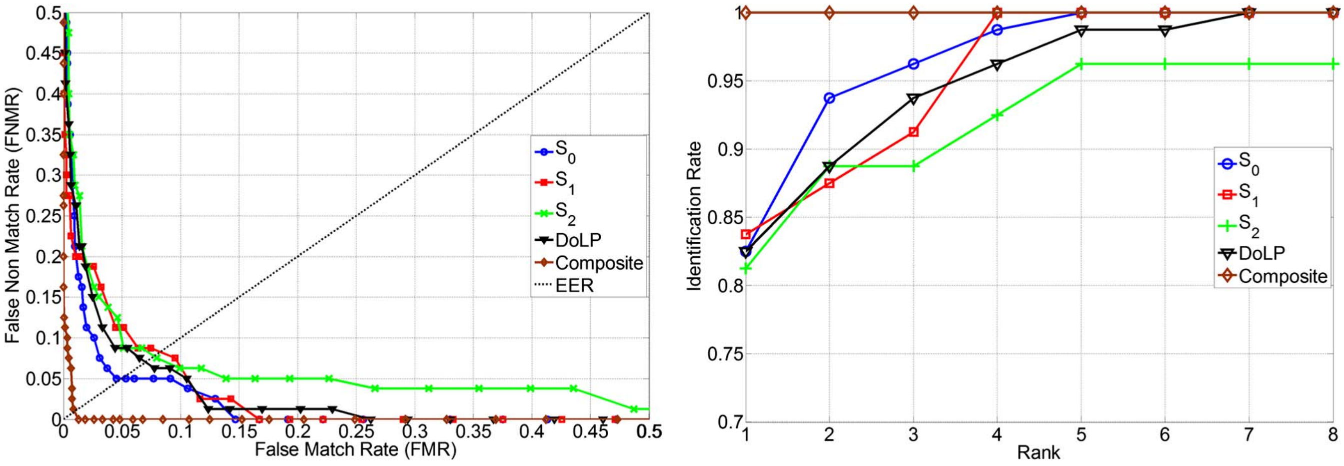


Fig. 3. Composite HOG representation where red, green, and blue color channel intensities correspond to S_0, S_1, and S_2 component contributions, respectively.

level of correlation to the visible spectrum feature representation, thus facilitating cross-modal face recognition. Here, the S_0, S_1, and S_2 components are represented by red, green, and blue colors, respectively. The intensity of each color band indicates the strength of the local edges as measured in the corresponding Stokes image. Of note, the vectors colored green, blue, or a mixture of green-blue, indicate edge details that are not prominent in the conventional LWIR face signature.

We evaluate the performance of polarimetric LWIR for face recognition by using aggregate- and rank-based statistics, which are commonly used in the biometrics community. For user verification or authentication applications, receiver operator characteristic (ROC) curves summarize the performance of a system in terms of the false-positive rate (also referred to as the false-match rate in biometrics) with respect to the false-negative rate (referred to as the false-non-match rate in biometrics) [10]. For identification applications, a cumulative match characteristic (CMC) curve is used to describe the ability of the system to correctly identify an unknown probe sample from a set of known gallery images. In this scenario, the identification rate is reported as a function of rank, where the rank corresponds to the position of the true match in the score-based sorting of returned matches. Figure 4(a) shows ROC curves and Fig. 4(b) shows CMC curves when using S_0, S_1, S_2, DoLP, or composite representation as the probe set for cross-modal face recognition. The composite feature representation, which



(a) (b)

Fig. 4. (a) Aggregate and (b) rank-based statistical performance curves for cross-modal matching of traditional LWIR (S_0), individual stokes images (S_1 and S_2), DoLP, and proposed Composite.

Table 1. Key ROC and CMC Data Points from Testing on 20 Subject Dataset

Input	EER (%)	FMR100 (%)	FMR1000 (%)	Rank-1 ID (%)
S_0	5.0	21.0	70.0	82.5
S_1	7.6	20.4	38.0	83.7
S_2	7.8	28.7	70.7	81.2
DoLP	7.0	26.6	47.7	82.5
Composite	0.9	0.9	11.8	100

is derived from a combination of the Stokes images, yields the highest performance in terms of the lowest FNMR at all examined FMR in Fig. 4(a), compared to conventional thermal (S_0) or any individual Stokes components. Furthermore, the CMC curve shows that the composite feature representation yields the best Rank-1 identification rate of 100%, compared to conventional thermal, which has a Rank-1 identification rate of 82.5%.

Table 1 lists values for several common face recognition performance metrics from ROC and CMC curves. The Equal Error Rate (EER) represents the rate at which the false-positive and false-negative rate is equal. The FMR100 is the FNMR where the FMR is fixed at 1% and is a common measure for verification systems, as this scenario typically operates in a state where denying access to a genuine user is more acceptable than allowing access to a non-genuine user. The FMR1000 is the FNMR with the FMR fixed at 0.1%. The Rank-1 ID is derived from the CMC curve and indicates the identification rate of the system when considering only the top match.

As can be observed in Table 1, the composite feature set outperforms all other probe face representations across all performance metrics for cross-modal face recognition. The identification rate, when using the composite polarimetric information, is approximately 18% higher than those achieved by using conventional LWIR (S_0) or S_1 and S_2 individually, indicating that polarimetric information complements conventional thermal features. Therefore, polarimetric LWIR is more effective than conventional thermal imaging for cross-modal face recognition, for both authentication as well as identification applications. In conclusion, polarimetric LWIR holds significant potential for personnel identification in nighttime operations.

References

1. B. F. Klare and A. K. Jain, IEEE Trans. Pattern Anal. Mach. Intell. **35**, 1410 (2013).
2. J. Choi, S. Hu, S. S. Young, and L. S. Davis, Proc. SPIE DSS **8371**, 83711L (2012).
3. T. Bourlai, A. Ross, C. Chen, and L. Hornak, Proc. SPIE DSS **8371**, 83711K (2012).
4. K. P. Gurton, A. J. Yuffa, and G. W. Videen, Opt. Lett. **39**, 3857 (2014).
5. D. L. Jordan and G. Lewis, Opt. Lett. **19**, 692 (1994).
6. M. Born and E. Wolf, *Principles of Optics* (Pergamon, 1959).
7. N. Dalal and B. Triggs, Proc. IEEE CVPR **2**, 886 (2005).
8. C. Cortes and V. Vapnik, Mach. Learn. **20**, 273 (1995).
9. B. Scholkopf and A. J. Smola, *Learning with Kernels* (MIT, 2002).
10. A. K. Jain, P. Flynn, and A. Ross, eds. *Handbook of Biometrics* (Springer, 2008).