ICME GRAND CHALLENGE RESULTS ON HETEROGENEOUS FACE RECOGNITION: POLARIMETRIC THERMAL-TO-VISIBLE MATCHING

Benjamin S. Riggan[1,*], *Nathaniel J. Short*[1,2], *M. Saquib Sarfraz*[3], *Shuowen Hu*[1]

[1] U.S. Army Research Laboratory, 2800 Powder Mill Rd., Adelphi, MD 20783
[2] Booz Allen Hamilton, 8283 Greensboro Dr., McLean, VA 22102
[3] Institute of Anthropomatics & Robotics, Karlsruhe Institute of Technology (KIT), Karlsruhe, Germany

He Zhang, Vishal M. Patel

Rutgers University Team

Sanka Rasnayaka , Jing Li, Terence Sim

National University of Singapore Team

Seyed Mehdi Iranmanesh, Nasser M. Nasrabadi

West Virginia University Team

ABSTRACT

This paper describes the IEEE ICME Grand Challenge on Heterogeneous Face Recognition (Polarimetric Thermal to Visible Matching), presents the submitted face recognition algorithms, and details the evaluation results. The challenge problem, sponsored by ICME and Polaris Sensor Technologies, is motivated by nighttime face recognition and compares state-of-the-art domain adaptive algorithms for cross-spectrum face recognition. Using unique databases containing corresponding polarimetric thermal and visible facial imagery, the algorithms were developed and independently evaluated. A brief summary of each algorithm is described, and the face verification performances in term of equal error rate (EER) and area under the curve (AUC) are reported. The best performing algorithm was a GAN-based approach submitted by the Rutgers University Team.

***Index Terms*—** heterogeneous face recognition

1. INTRODUCTION

Facial biometrics are becoming increasingly popular and effective for use in security, surveillance, and forensic based applications. While deep learning based approaches have demonstrated robust performance for face recognition using imagery acquired in the visible spectrum, there has been significantly less research on the topic of heterogeneous face recognition, especially related to matching thermal (i.e., midwave and longwave) infrared and visible faces. Thermal imagery, which does not require active illumination since thermal radiation is naturally emitted from faces, is ideal for nighttime face recognition. The primary challenge with acquiring thermal imagery for face recognition is the lack of databases/watchlists containing thermal facial signatures. Therefore, a thermal probe image must be matched against visible imagery within an existing biometric database.

Hu et al. [1] proposed a one-versus-all framework using partial least squares (PLS) classifiers to facilitate thermal-to-visible face recognition. In this framework, they use a few positive exemplars imagery and many negative exemplars from visible spectrum imagery to build a classifier for each person. For added classifier robustness in the thermal-to-visible face recognition task, negative thermal samples are used to augment the negative exemplars. In [2] a deep perceptual mapping (DPM), which maps visible to thermal representations, demonstrated improved results.

More recently, polarimetric thermal imaging has been used to enhance cross-spectrum face recognition performance. Polarimetric thermal imagery is represented by Stokes parameters: S_0, S_1, S_2, where S_0 is conventional thermal image, S_1 captures the differences between the the 0 degree and 90 degree polarization states, and S_2 captures the difference between the 45 degree and 135 degree polarization states. From the Stokes representation the Degree of Linear Polarization can be computed as $DoLP = \sqrt{S_1^2 + S_2^2}/S_0$. Short et al. [3, 4] proposed the use of polarization state information for improving performance for nighttime face recognition by using a composite image created by a uniformly weighted combination of Stokes representations.

The objective of this grand challenge is for participants to submit state-of-the-art approaches for polarimetric thermal-

*Contact corresponding author at benjamin.s.riggan.civ@mail.mil

to-visible matching for independent testing and evaluation. The datasets and protocol for this challenge are described in Section 2.

2. DATABASES AND EVALUATION PROTOCOL

The ARL Polarimetric Thermal Face Database [5], which contains polarimetric thermal and visible imagery from 60 subjects at three different physical distances (2.5m, 5m, and 7.5m) with neutral (baseline) and variable expression conditions, is provided to each participant for algorithm training. Further details about this dataset can be found in [5]. ARL also collected a more recent dataset at a different location, which served as the sequestered dataset for evaluation of the submitted algorithms. This dataset contains corresponding visible and polarimetric thermal imagery from 51 subjects collected at 2.5m with baseline and expression conditions. Notable differences, other than time and location of two collections, include: relative sensor placement (i.e., differences in field of views), subject diversity (e.g., race, sex, gender), and ambient temperature. These differences introduce some degree of variability between the two datasets that test the robustness of the submitted algorithms.

For this challenge, we independently evaluate each submitted algorithm using the face verification protocol, which takes a pair of images and returns a match score indicating the degree of similarity between the two images. The key difference in this challenge compared to a typical face verification experiment is that the image pairs do not come from the same imaging modality. Instead, a match score is returned for a polarimetric thermal and visible image pair. For this challenge, we use the 51 subject collection to evaluate submitted algorithms, which are trained using the ARL Polarimetric Thermal Database. Using the 51 subject collection, we generated 1500 polarimetric thermal and visible image pairs that are comprised of 500 matching pairs and 1000 non-matching pairs for algorithm test and evaluation. There are approximately 10 matching pairs and 20 non-matching pairs per subject that cover a variety of facial expressions at a single range.

3. CHALLENGE SUBMISSIONS

Three teams: Rutgers University, National University of Singapore (NUS), West Virginia University (WVU) submitted algorithms to this ICME heterogenous face recognition challenge. In this section, we provide a brief overview of each approach.

The Rutgers University Team proposed a Generative Adversarial Network (GAN) based multi-stream feature-level fusion technique to synthesize high-quality visible images from polarimetric thermal images. This type of synthesis based approach is motivated by initial work in [6]. The proposed GAN-based network consists of a generator, a discriminator sub-network and a deep guided sub-network (see

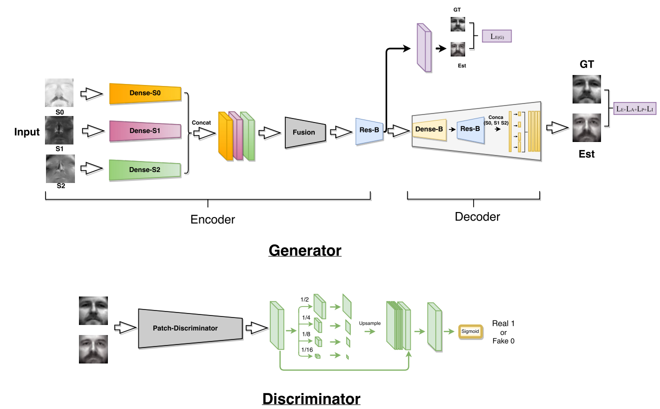

Fig. 1: An overview of the proposed GAN-based multi-stream encoder-decoder network from Rutgers Univerisy. The generator contains a multi-stream feature-level fusion encoder-decoder network. In addition, a deep-guided subnet is stacked at the end of the encoding part. The discriminator is composed of a multi-scale patch-discriminator structure.

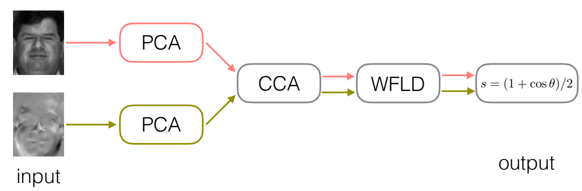

Fig. 2: Proposed framework from NUS team that finds correlation between visible (top) and S_2 (bottom) images.

Fig. 1). The generator is composed of a multi-stream encoder-decoder network based on dense-residual blocks [7, 8, 9], the discriminator is designed to capture features at multiple-scales for discrimination and the deep guided sub-net aims to guarantee that the encoded features contain geometric and texture information to recover the visible face. To further enhance the network's performance, it is guided by perceptual loss and an identity preserving loss [10] in addition to adversarial loss. Once the face images are synthesized, the vgg-face network, pre-trained with only visible face imagery, is used for matching.

For this challenge, vgg-face features from visible and synthesized images are extracted. Using these deep features, the cosine distance in (1) is used to produce the match score.

$$s = (1 + \cos\theta)/2 = (1 + \frac{\boldsymbol{v}_v \boldsymbol{v}_t}{\|\boldsymbol{v}_v\| \|\boldsymbol{v}_t\|})/2 \qquad (1)$$

The National University of Singapore Team proposed a simple but computationally efficient method for thermal-to-visible face verification, as shown in **Fig. 2**. To overcome the modality gap between polarimetric thermal and visible imagery, they first apply Principle Component Analysis (PCA) to extract the most prominent features from each modality. Then, they associate these features by employing Canonical

Correlation Analysis (CCA) [11] to project the the two types of images into a unified feature space. By maximizing the cross-covariance across two types of images, CCA captures the crucial mutual information of visible and thermal images for face verification. Subsequently, they apply a Whitened Fisher Linear Discriminant (WFLD)[12] analysis on the vectors in unified feature space so that their identities are distinguished. Finally, they compute the similarity of a thermal face image and a visible face image using the cosine distance of their WFLD feature vectors. In other words, the similarity score s is obtained using (1), where $\boldsymbol{v}_v$ and $\boldsymbol{v}_t$ are the WFLD feature vectors of visible image and thermal image respectively.

The West Virginia University Team proposed a vgg-16 like network [13] for thermal-to-visible face recognition. The vgg-16 neural network is comprised of five major convolutional components which are connected in series. The main difference between the proposed and vgg-16 networks is in the last component, where global pooling is used instead of the max pooling to reduce the number of parameters.

Two vgg-16 like networks: a visible spectrum (Vis-DCNN) and a polarimetric thermal (Pol-DCNN) are coupled (similar to a Siamese network [14] without weight sharing) to find the latent deep features representing the common relationship between the polarimetric thermal face images and their corresponding visible ones. These two networks are trained via a contrastive loss function [14], which minimizes the distance between corresponding image pairs and maximizes the distance between non-corresponding image pairs in the latent embedding subspace. An overview of the model is shown in **Fig. 3**.

Similar to [14], the contrastive loss is of the form of:

$$\begin{aligned} \ell_{cont}(z_1(x_i^{vis}), z_2(x_j^{pol}), y_{cont}) = \\ (1 - y_{cont}) L_{gen}(D(z_1(x_i^{vis}), z_2(x_j^{pol})) + \\ y_{cont} L_{imp}(D(z_1(x_i^{vis}), z_2(x_j^{pol})) \ , \end{aligned} \tag{2}$$

where x_i^{vis} is the visible face image, x_j^{pol} is the polarimetric face images, y_{cont} is a binary label, L_{gen} and L_{imp} represent the partial loss functions for the genuine and impostor pairs, respectively, and $D(z_1(x_i^{vis}), z_2(x_j^{pol}))$ indicates the Euclidean distance between the embedded data in the common feature subspace. The binary label, y_{cont}, is assigned a value of 0 for a genuine image pair. On the contrary, when the inputs are from different classes (i.e., an impostor pair), y_{cont} is equal to 1. L_{gen} and L_{imp} are defined as follows:

$$L_{gen}(D(z_1(x_i^{vis}), z_2(x_j^{pol}))) = \frac{1}{2} D(z_1(x_i^{vis}), z_2(x_j^{pol}))^2 \tag{3}$$

$$\begin{aligned} L_{imp}(D(z_1(x_i^{vis}), z_2(x_j^{pol}))) = \\ \frac{1}{2} max(0, m - D(z_1(x_i^{vis}), z_2(x_j^{pol})))^2 \quad \text{for} \quad y_i \neq y_j \ . \end{aligned} \tag{4}$$

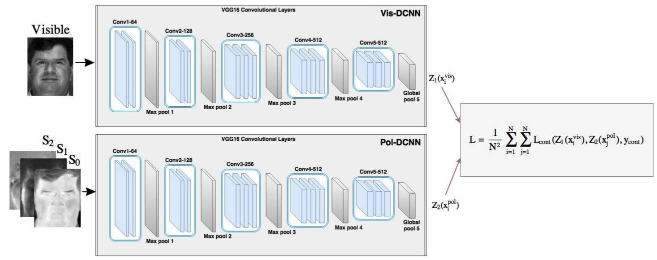

Fig. 3: Proposed network submitted by WVU using two convolutional networks (Vis-DCNN and Pol-DCNN) coupled by contrastive loss function.

After training, the deep coupled network model transforms visible and polarimetric face images into a common discriminative embedding space. After mapping visible and polarimetric face images, the Euclidean distance is reported a dissimilarity metric for face verification.

4. CHALLENGE RESULTS

For each algorithm submission (Section 3) and baseline algorithm, we compute the set of similiary (or dissimilarity) scores for matching and non-matching pairs. From the set of scores, we generated receiver operating characteristic (ROC) curves where the equal error rate (EER) and area under the curve AUC are used for purposes of comparison.

The baseline algorithm is similar to the coupled neural network (CpNN) method described in [15], except the similarity score is provided by the cosine similarity measure (opposed to a discriminative PLS classification score).

Fig. 4 shows the ROC curves for each algorithm, including the baseline. The plot shows that the Rutgers University Team achieved the best verification performance with an EER of 5.77% and AUC of 98.63%. The West Virginia University Team achieved an EER of 31.01% and AUC of 75.15%, which is similar to the baseline algorithm. Lastly, the National University of Singapore Team achieved an EER of 37.97% and AUC of 67.93%, which is worse than the baseline.

The mean run times are 270.9, 260.0, and 2.74 seconds for Rutgers, WVU, and NUS algorithms, respectively.

5. CONCLUSIONS

This challenge on polarimetric thermal-to-visible matching provided independent testing and evaluation on a 51 subject dataset that was collected at different time and location from the training data. Moreover, the 51 subject dataset has no subjects in common with the dataset used by participants to train their algorithms. This challenge setup provides a significantly challenging task, which is motivated by the need to develop advanced technology for nighttime face recognition.

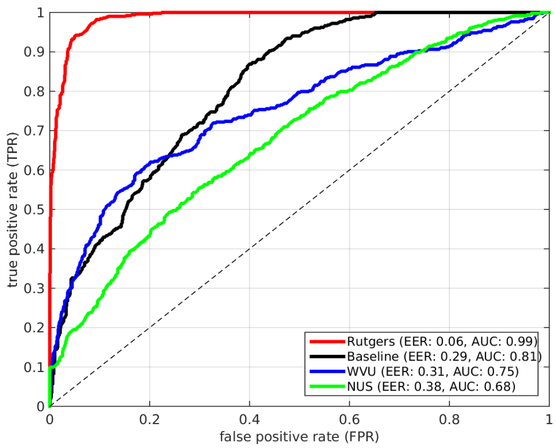

Fig. 4: Comparison of ROC curves.

Based on performer algorithms and the evaluation protocol for this challenge, the state-of-the-art performance for polarimetric thermal-to-visible face verification is an EER of 5.77% and AUC of 98.63%. Considering systems do not generally operate at the threshold corresponding to the EER for security related applications, the state-of-the-art algorithm achieved approximately 84% and 67% true positive rates at 3% and 1% false positive rates, respectively. To our knowledge, this is the first challenge on polarimetric thermal-to-visible face recognition. The excellent results achieved by the Rutgers University team and good results by the West Virginia University team demonstrate current state-of-the-art and promising future of this technology for nighttime face recognition.

6. REFERENCES

[1] Shuowen Hu, Jonghyun Choi, Alex L Chan, and William Robson Schwartz, "Thermal-to-visible face recognition using partial least squares," *JOSA A*, vol. 32, no. 3, pp. 431–442, 2015.

[2] M Saquib Sarfraz and Rainer Stiefelhagen, "Deep perceptual mapping for cross-modal face recognition," *International Journal of Computer Vision*, vol. 122, no. 3, pp. 426–438, 2017.

[3] Nathaniel Short, Shuowen Hu, Prudhvi Gurram, Kristan Gurton, and Alex Chan, "Improving cross-modal face recognition using polarimetric imaging," *Optics letters*, vol. 40, no. 6, pp. 882–885, 2015.

[4] Nathaniel Short, Shuowen Hu, Prudhvi Gurram, and Kristan Gurton, "Exploiting polarization-state information for cross-spectrum face recognition," in *Biometrics Theory, Applications and Systems (BTAS), 2015 IEEE 7th International Conference on*. IEEE, 2015, pp. 1–6.

[5] Shuowen Hu, Nathaniel J Short, Benjamin S Riggan, Christopher Gordon, Kristan P Gurton, Matthew Thielke, Prudhvi Gurram, and Alex L Chan, "A polarimetric thermal database for face recognition research," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops*, 2016, pp. 119–126.

[6] Benjamin S Riggan, Nathanial J Short, Shuowen Hu, and Heesung Kwon, "Estimation of visible spectrum faces from polarimetric thermal faces," in *Biometrics Theory, Applications and Systems (BTAS), 2016 IEEE 8th International Conference on*. IEEE, 2016, pp. 1–7.

[7] Kaiming He, Xiangyu Zhang, Shaoqing Ren, and Jian Sun, "Deep residual learning for image recognition," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2016, pp. 770–778.

[8] Gao Huang, Zhuang Liu, Kilian Q Weinberger, and Laurens van der Maaten, "Densely connected convolutional networks," *arXiv preprint arXiv:1608.06993*, 2016.

[9] He Zhang and Vishal M Patel, "Densely connected pyramid dehazing network," in *The IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, June 2018.

[10] He Zhang, Vishal M Patel, Benjamin S Riggan, and Shuowen Hu, "Generative adversarial network-based synthesis of visible faces from polarimetric thermal faces," *International Joint Conference on Biometrics 2017*, 2017.

[11] David R Hardoon, Sandor R Szedmak, and John R Shawe-taylor, "Canonical correlation analysis: An overview with application to learning methods," *Neural Comput.*, vol. 16, no. 12, pp. 2639–2664, Dec. 2004.

[12] Sheng Zhang, Terrance Sim, and Mei-Chen Yeh, "Identity and variation spaces: Revisiting the fisher linear discriminant," in *ICCVW*, 2009.

[13] Karen Simonyan and Andrew Zisserman, "Very deep convolutional networks for large-scale image recognition," *ICLR*, 2014.

[14] Sumit Chopra, Raia Hadsell, and Yann LeCun, "Learning a similarity metric discriminatively, with application to face verification," in *Computer Vision and Pattern Recognition, 2005. CVPR 2005. IEEE Computer Society Conference on*. IEEE, 2005, vol. 1, pp. 539–546.

[15] Benjamin S Riggan, Nathaniel J Short, and Shuowen Hu, "Optimal feature learning and discriminative framework for polarimetric thermal to visible face recognition," in *Applications of Computer Vision (WACV), 2016 IEEE Winter Conference on*. IEEE, 2016, pp. 1–7.