Generative Adversarial Network-based Synthesis of Visible Faces from Polarimetric Thermal Faces

He Zhang[1], Vishal M. Patel[1], Benjamin S. Riggan[2], and Shuowen Hu[2]

[1] Rutgers, The State University of New Jersey, Piscataway, NJ 08854
[2] U.S. Army Research Laboratory, 2800 Powder Mill Rd., Adelphi, MD 20783

{he.zhang92,vishal.m.patel}@rutgers.edu, {benjamin.s.riggan.civ, shuowen.hu}@mail.mil

Abstract

The large domain discrepancy between faces captured in polarimetric (or conventional) thermal and visible domain makes cross-domain face recognition quite a challenging problem for both human-examiners and computer vision algorithms. Previous approaches utilize a two-step procedure (visible feature estimation and visible image reconstruction) to synthesize the visible image given the corresponding polarimetric thermal image. However, these are regarded as two disjoint steps and hence may hinder the performance of visible face reconstruction. We argue that joint optimization would be a better way to reconstruct more photo-realistic images for both computer vision algorithms and human-examiners to examine. To this end, this paper proposes a Generative Adversarial Network-based Visible Face Synthesis (GAN-VFS) method to synthesize more photo-realistic visible face images from their corresponding polarimetric images. To ensure that the encoded visible-features contain more semantically meaningful information in reconstructing the visible face image, a guidance sub-network is involved into the training procedure. To achieve photo realistic property while preserving discriminative characteristics for the reconstructed outputs, an identity loss combined with the perceptual loss are optimized in the framework. Multiple experiments evaluated on different experimental protocols demonstrate that the proposed method achieves state-of-the-art performance.

1. Introduction

Face recognition is one of the most widely studied topics in computer vision. In recent years, methods based on deep convolutional neural networks (CNNs) have shown impressive performance improvements for face recognition problems [17, 22, 26]. Even though these methods are able to address many challenges such as the low-resolution,
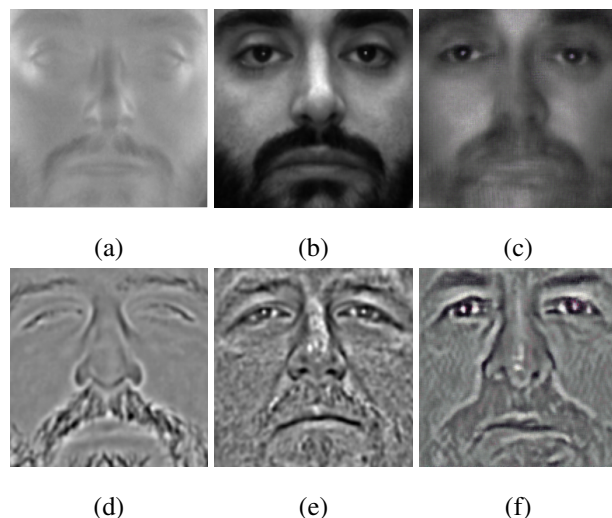

Figure 1: Sample results of our proposed method, GAN-VSF. (a) Input Polarimetric image. (b) Ground truth visible image. (c) Estimated visible image using GAN-VSF from the image shown in (a). (d), (e), and (f) are the DoG filtered images corresponding to the images shown in (a), (b), and (c), respectively.

pose variation and illumination variation to some extent, they are specifically designed for recognizing face images that are collected near-visible spectrum. They often do not perform well on the face images captured from other domains such as polarimetric [6, 20, 23], infrared [11, 16] or millimeter wave [3] due to significant phenomenological differences as well as a lack of sufficient training data. Distributional change between thermal and visible images makes thermal-to-visible face recognition very challenging. Various methods have been developed in the literature for bridging this gap, seeking to develop a cross-domain face recognition algorithm [21, 1, 13, 24]. In particular, recent works have proposed using the polarization-state informa-

tion of thermal emissions to enhance the performance of cross-spectrum face recognition [6, 20, 23]. It has been shown that polarimetric-thermal images capture geometric and textural details of faces that are not present in the conventional thermal facial imagery [23]. As a result, the use of polarization-state information can perform better than using only conventional thermal imaging for face recognition.

Previous approaches have attempted to improve the cross-domain face recognition performance, however, it is still of great importance to guarantee that the human examiners can identify whether the given polarimetric and the visible image share the same identity or not. Consider the polarimetric face image shown in Figure 1(a). The corresponding visible image is shown in Figure 1(b). As can be seen from these images, it is extremely difficult for either human-examiners or existing face recognition systems to determine whether these two images share the same identity. Hence, methods that can automatically generate high quality visible images from their corresponding polarimetric images are needed.

One such method was recently proposed in [20]. However, it suffers from the following drawbacks. 1) The procedures of visible feature extraction and image reconstruction are not jointly optimized and hence may degrade the reconstruction performance. 2) The recovered images are not photo realistic and hence may hinder the verification performance. In this work, we take a major step towards addressing these two issues. In order to optimize the overall algorithm jointly, a generative adversarial network (GAN) based approach with a guidance sub-network is proposed to generate more sharp visible faces. Furthermore, to generate more photo realistic images and to make sure that the generated images contain highly discriminative information, an identity loss combined with the perceptual loss is included in the optimization. Quantitative and qualitative experiments evaluated on four different polarimetric-to-visible settings demonstrate that the proposed method can achieve state-of-the-art performance as compared to previous methods.

Samples results of the proposed Generative Adversarial Network-based Visible Face Synthesis (GAN-VFS) method are shown in Figure 1. Given the polarimetric face images shown in Figure 1(a), our method automatically synthesizes a visible image shown in Figure 1(c), which is very close to the ground truth visible image shown in Figure 1(b). Images in Figure 1(d)-(f) show the difference-of-Gaussian (DoG) filtered images corresponding to the images shown in Figure 1(a)-(c).

2. Generative Adversarial Networks (GANs)

Generative Adversarial Networks were proposed by Goodfellow et al. in [4] to synthesize realistic images by effectively learning the distribution of training images. The authors adopted a game theoretic min-max optimization

framework to simultaneously train two models: a generative model, G, and a discriminative model, D. The success of GANs in synthesizing realistic images has led to researchers exploring the adversarial loss for numerous low-level vision applications such as style transfer [14], image in-painting [18], image to image translation [19], image super-resolution [12] and image de-raining [29]. Inspired by the success of these methods, we propose to use the adversarial loss to learn the distribution of visible face images for their accurate estimation.

3. Proposed Method

Instead of optimizing the two procedures (visible feature estimation and visible image reconstruction) separately, a new unified GAN-based framework is proposed in this section. In the following sub-sections, we discuss these important parts in detail starting with the GAN objective function followed by details of the proposed network and the loss functions.

3.1. GAN Objective Function

In order to learn a good generator G so as to fool the learned discriminator D and to make the discriminator D good enough to distinguish synthesized visible image from real ground truth, the proposed method alternatively updates G and D following the structure proposed in [8, 29]. Given an input polarimetric image X, conditional GAN aims to learn a mapping function to generate output image Y by solving the following optimization problem:

$$\min_G \max_D \quad \mathbb{E}_{X \sim p_{data}(X)}[\log(1 - D(X, G(X)))] + \\ \mathbb{E}_{X \sim p_{data}(X,Y)}[\log D(X, Y)]. \quad (1)$$

3.2. Network Overview

The process of transforming the image from polarimetric domain to visible domain can be regarded as a pixel-level image regression problem. Basically, the transforming procedure can be divided into two separated processes as discussed in [20]. Firstly, a set of features that can aid the process of reconstructing visible image are extracted from the given polarimetric images. Then, an optimization procedure is proposed to reconstruct the corresponding visible face images via those learned visible features. Even though the previous method [20] achieves very good performance in transforming the image from polarimetric domain to visible domain via the proposed two steps, the proposed two steps are not jointly learned and optimized. In other words, the proposed two-step method has to rely on the assumption that the learned features contain semantically meaningful information in reconstructing the visible face images.

To overcome this issue and to relax that assumption, a unified GAN-based synthesis network that can directly
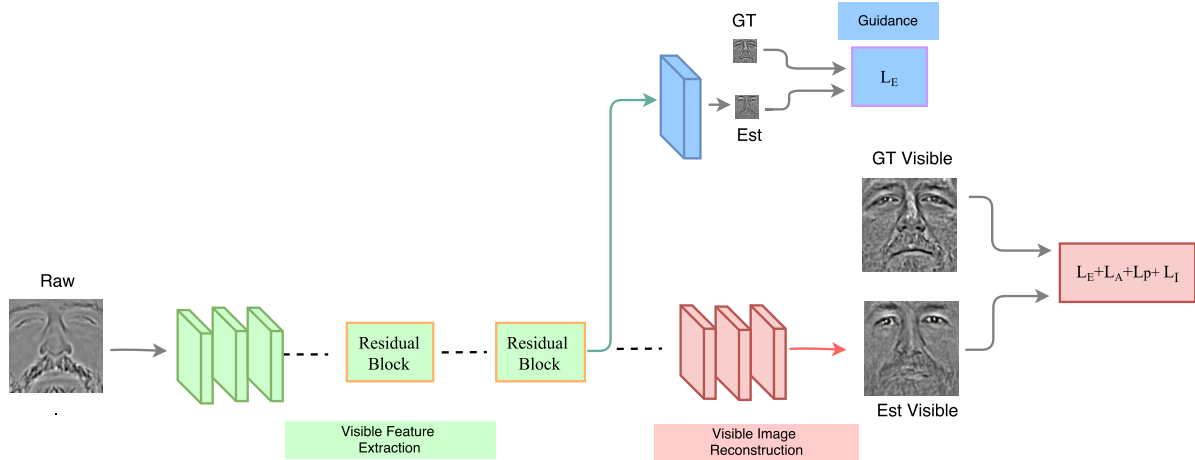
Figure 2: Overview of the proposed GAN-VFS method. It contains three modules. (a) Visible feature extraction module, (b) Guidance sub-network and (c) Visible image reconstruction module. Firstly, the visible feature is extracted from the raw polarimetric image. Then, to make sure that the learned feature can better reconstruct the visible image, a guidance sub-network is involved into the optimization. Finally, the guided feature is used to reconstruct the photo realistic visible image using the combination of different losses.

learn an end-to-end mapping between the polarimetric image and its corresponding visible image is proposed. The whole network contains an encoder-decoder structure, where the learned visible features can be regarded as the outputs of the encoder part and input for the decoder part. To guarantee the reconstructability of the encoded features and to make sure that the leaned features contain semantic information, a guidance sub-network [27], is introduced at the end of the visible feature extraction part. The overall network architecture is shown in Figure 2.

To overcome the side effect of blurry results brought by the traditional Euclidean loss (L_E loss) [1] and to discriminate the generated visible face images from their corresponding ground truth, a GAN structure is deployed. Even though GAN's structure can generate more reasonable results compared to the tradition L_E loss, as will be shown later, the results generated by the traditional GAN contain undesirable facial artifacts, resulting in a less photo realistic image. To address this issue and meanwhile generate visually pleasing results, the perceptual loss is included in our work, where the perceptual loss is evaluated on pre-trained VGG-16 models, as discussed in [9, 28].

As the ultimate goal of the our proposed synthesis method is to guarantee that human examiners can identify the person given his synthesized face images, it is also important to involve the discriminative information into consideration. Similar to the perceptual loss, we propose an identity loss that is evaluated on a certain layer of the fine-tuned VGG-Polar model. The VGG-Polar model is fine-tuned using the visible images with its corresponding labels

from the Polarimetric-Visible dataset [6].

3.3. Network Structure

Figure 2 gives an overview of the proposed synthesis framework. We adopt an encoder-decoder structure as the basis in the generator part. Basically, the proposed generator can be divided into two parts. Firstly, a set of convolutional layers with stride 2 combined with a set of residual blocks [9] are regarded as visible feature estimation part. Specifically, the residual blocks are composed of two convolutional layers with 3×3 kernels and 64 feature maps followed by batch-normalization layers [7] and PReLU [5] as the activation function. Then, a set of transpose convolutional layers with stride 2 are denoted as the visible image reconstruction procedure. To make sure that the transformed features contain enough semantic information, a guided sub-part is enforced in the network. Meanwhile, to make the generated visible face images indistinguishable from the ground truth visible face images, a CNN-based differentiable discriminator is used as a guidance to guide the generator in generating better visual results. For the discriminator, we use PatchGANs [8] to discriminate whether the given images are real or fake.

The structure of the proposed discriminator network is as follows:
CB(32)-CBP(64)-CBP(128)-CBP(256)-CBP(256)-C(1)-Sigmoid,

where C represents the covolutional layer, P indicates Prelu [5] and B denotes as batch-normalization [7]. The number in the bracket represents the number of output feature maps

---

[1]We use L_E loss to represent the Euclidean loss throughout the paper.

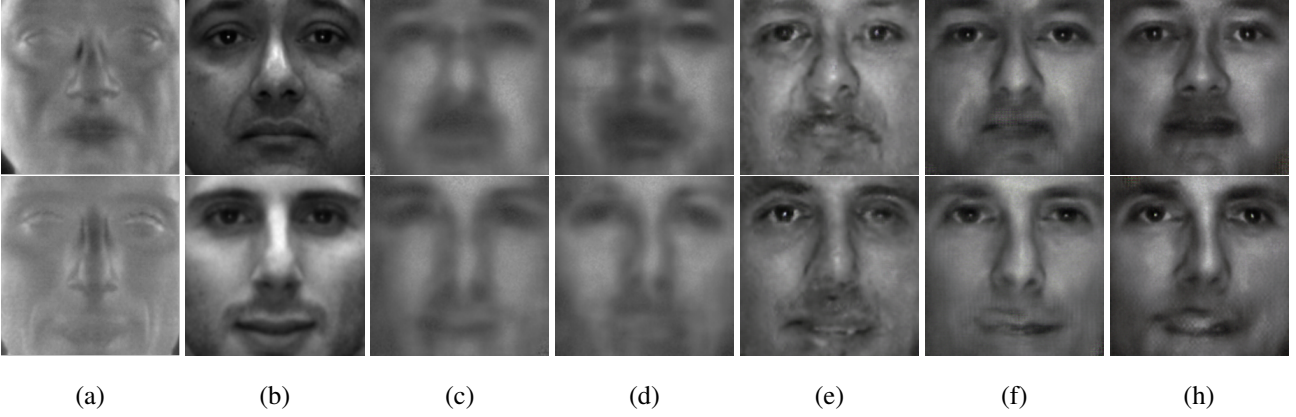| | | | | | | |
| (a) | (b) | (c) | (d) | (e) | (f) | (h) |

Figure 3: Comparisons of visible face images synthesized with different experimental configurations. (a) Raw polarimetric images. (b) Ground truth visible images. (c) Reconstructed images with only L_E loss. (d) Reconstructed images with L_E loss and guidance L_E loss. (e) Reconstructed images with GAN, L_E and guidance L_E loses. (f) Reconstructed images with GAN, L_E, guidance L_E and perceptual losses. (h) Reconstructed images using with all five losses.

of the corresponding layer.

3.4. Loss Functions

The proposed method contains the following loss functions: the Euclidean $L_{E(G)}$ loss enforced on the recovered visible image, the L_E loss enforced on the guidance part, the adversarial loss to guarantee more sharp results, the perceptual loss to preserve more photo realistic details and the identity loss to preserve more distinguishable information for the outputs. The overall loss function is defined as follows

$$L_{\text{all}} = L_E + L_{E(G)} + \lambda_A L_A + \lambda_P L_P + \lambda_I L_I, \quad (2)$$

where L_E denotes the Euclidean loss, $L_{E(G)}$ denotes the Euclidean loss on the guidance sub-network, L_A represents the adversarial loss, L_P indicates the perceptual loss and L_I is the identity loss. Here, λ_A, λ_P and λ_I are the corresponding weights.

The L_E and the adversarial losses are defined as follows:

$$L_E, L_{E(G)} = \frac{1}{WH} \sum_{w=1}^{W} \sum_{h=1}^{H} \|\phi_G(I)^{w,h} - Y_t^{w,h}\|_2, \quad (3)$$

$$L_A = -\log(\phi_D(\phi_G(I)^{c,w,h})), \quad (4)$$

where I is the input polarimetric image, Y_t is the ground truth visible image, $W \times H$ is the dimension of the input image, ϕ_G is the generator sub-network G and ϕ_D is the discriminator sub-network D.

As the perceptual loss and the identity losses are evaluated on a certain layer of the given CNN model, both can be

defined as follows:

$$L_{P,I} = \frac{1}{C_i W_i H_i} \sum_{c=1}^{C_i} \sum_{w=1}^{W_i} \sum_{h=1}^{H_i} \|V(\phi_E(I))^{c,w,h} - V(Y_t)^{c,w,h}\|_2, \quad (5)$$

where Y_t is the ground truth visible image, ϕ_E is the proposed generator, V represents a non-linear CNN transformation and C_i, W_i, H_i are the dimensions of a certain high level layer V, which differs for perceptual and identity losses.

3.5. Implementation

The entire network is trained on a Nvidia Titan-X GPU using the Torch-7 framework [2]. We choose $\lambda_A = 0.005$ for the adversarial loss, $\lambda_P = 0.8$ for the perceptual loss and $\lambda_I = 0.1$ for the identity loss. During training, we use ADAM [10] as the optimization algorithm with learning rate of 8×10^{-4} and batch size of 3 images. All the preprocessed training samples are resized to 256×256. The perceptual loss is evaluated on relu3-1 layers the in the pretrained VGG [17] model. The identity loss is evaluated on the relu2-2 layer of the fine-tuned VGG-Polar model.

4. Experimental Results

In this section, we demonstrate the effectiveness of the proposed approach by conducting various experiments on a unique dataset that contains polarimetric and visible image pairs from 60 subjects [6]. Basically, a polarimetric image, referred as Stokes images is composed of three channels: S0, S1 and S2. Here, S0 represents the conventional thermal image, whereas S1 and S2 represent the horizontal/vertical and diagonal polarization-state information, respectively. The dataset was collected at three distances:

Range 1 (2.5 m), Range 2 (5 m), and Range 3 (7.5 m). Following the protocol defined in [20], we only use the data corresponding to range 1 (baseline and expression). In particular, 30 subjects from range 1 are selected for training and the remaining 30 subjects are used for evaluation. We repeat this process multiple times and report average results. We evaluate the performance of different methods on the tightly cropped DoG filtered images [20]. We also conduct another set of experiments to directly transfer the polarimetric images to visible images. To summarize, our proposed network is evaluated on the following four protocols:

(a) Conventional thermal (S0) to Visible (Vis). [2]
(b) Polarimetric thermal (Polar) to Visible (Vis).
(c) DoG of S0 (S0-DoG) to DoG of Visible (Vis-DoG).
(d) DoG of Polarimetric thermal (Polar-DoG) to DoG of Visible (Vis-DoG).

4.1. Ablation Study

In order to better demonstrate the improvements obtained by different modules in the proposed network, we perform an ablation study involving the following experiments: 1) Polar to Visible estimation with only L_E loss, 2) Polar to Visible estimation with L_E loss and guidance L_E loss, 3) Polar to Visible estimation with GAN, L_E and guidance L_E loss, (4) Polar to Visible estimation with GAN, L_E, guidance L_E loss and perceptual loss, and finally (5) Polar to Visible estimation with all five losses. Due to the space limitations, the ablation study is only evaluated on the polar-to-visible image synthesis experiments. Two reconstruction results are shown in Figure 3. It can be observed from the results that the L_E loss itself generates much blurry faces. Even though the results with L_E guidance sub-network are slightly better with more details, they are still very blured and many high frequency details are missing. By involving the GAN structure in the proposed methods, more details are added to the results. But it can be observed that GAN itself produces images with artifacts. The introduction of the perceptual loss in the proposed framework is able to tackle the artifacts better and makes the results visually pleasing. Finally, the combination of all five losses can generate more reasonable results.

To better demonstrate the effectiveness of the different losses in the proposed methods, we plot the receiver operating characteristic (ROC) curves corresponding to the discussed five different network settings. The results are shown in Figure 4. All the verification results are evaluated on the deep features extracted from the VGG-face model [17] without fine-tuning. From the ROC curves, it can be clearly observed that even though the identity loss does not produce visually different results, it can bring in more discriminative information.
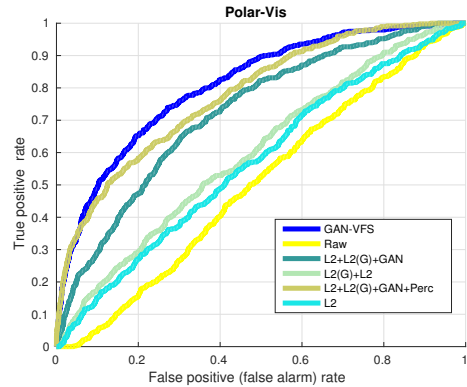

Figure 4: The ROC curves corresponding to the ablation study.

4.2. Comparison with State-of-art Methods

To demonstrate the improvements achieved by the proposed method, it is compared against recent state-of-the-art methods on the discussed datasets. In particular, we compare the performance of our method with that of the following two recent methods: Mahendran. *et al.* [15] and Riggan. *et al.* [20].

As discussed above, four sets of experiments are conducted in evaluating the performance both qualitatively and quantitatively. Peak signal-to-noise ratio (PSNR) and structural similarity (SSIM) [25] are used to evaluate the synthesized image quality performance of different methods.

Results corresponding to the proposed method and recent state-of-the-art methods on two sample images from the test dataset are shown in Figure 5. It can be observed from all four experiments (S0-Vis, Polar-Vis, S0-Vis (DoG) and Polar-Vis (DoG)) that the recovered results from our proposed method (last row) are more photo realistic and tend to recover more details. This can be clearly seen by comparing the eye regions of the recovered images.

The quantitative results evaluated using PSNR and SSIM are tabulated in Table 1 and Table 2, respectively. It can be clearly observed that the proposed method is able to achieve superior quantitative performance compared the previous approaches. These results highlight the significance of using a GAN-based approach to image synthesis.

Table 1: The average PSNR (dB) results corresponding to different methods.

	Mahendran. *et al.* [15]	Riggan. *et al.* [20]	**GAN-VFS**
S0-Vis	9.58	12.09	**17.11**
Poalr-Vis	9.85	12.54	**17.64**
S0-Vis (DoG)	11.05	15.02	**16.84**
Poalr-Vis (DoG)	11.72	15.64	**17.13**

---

[2] Conventional thermal (it is also referred as S0 in this paper.) measures the total intensity of thermal emissions.

| S0-Vis (DoG) | Polar-Vis (DoG) | S0-Vis | Polar-Vis |

Raw

Ground Truth
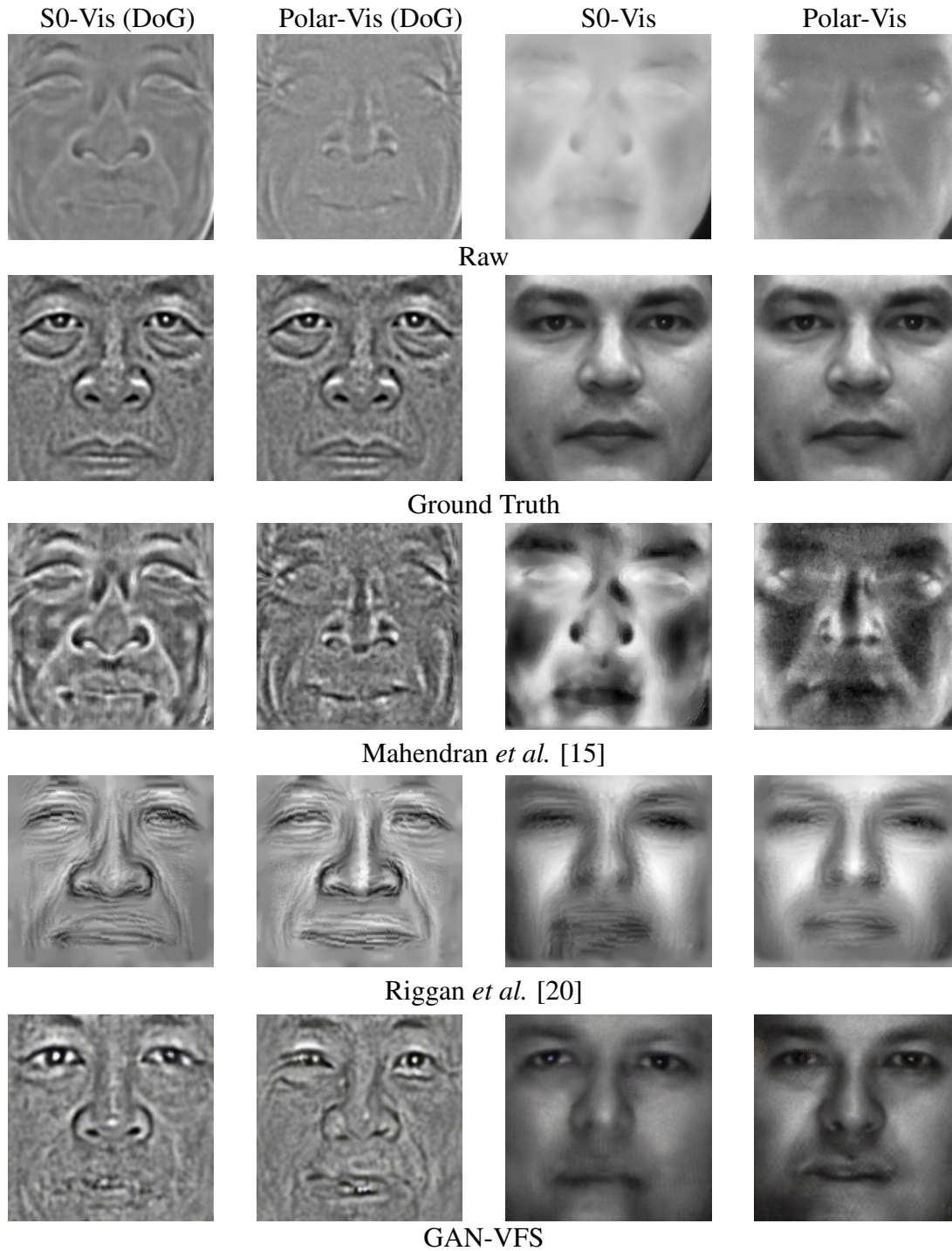
Mahendran *et al.* [15]

Riggan *et al.* [20]

GAN-VFS

Figure 5: Sample results corresponding to different methods on four protocols.

4.2.1 Face Verification Results

As the goal of synthesizing the visible image given the polarimetric image is to improve the verification results both for human examiners and computer vision algorithms, it is also important to evaluate how the recovered visible images work in face verification tasks. In this section, we propose to use the performance of face verification as a metric to evaluate the polarimetric-to-visible face synthesis algorithms. The experimental details are as follows: The ground truth images from all the 30 testing subjects are regarded as
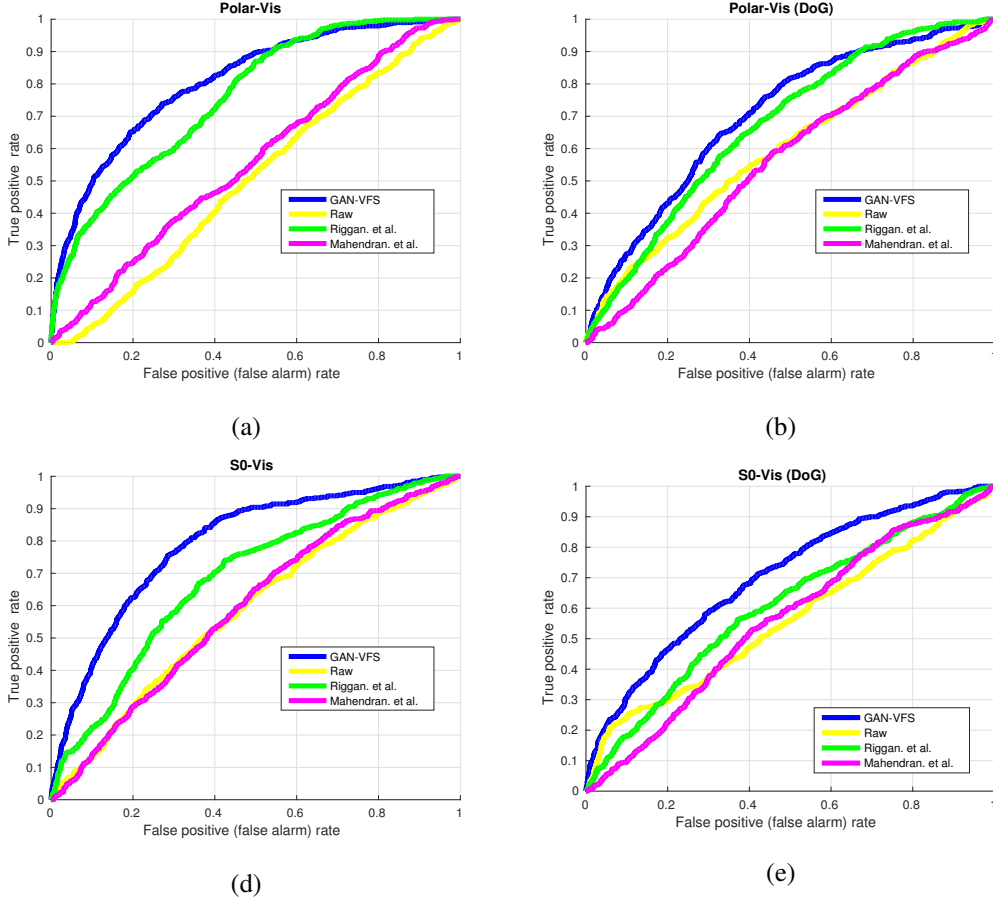
Figure 6: The verification results corresponding to different methods in terms of the ROC curves. (a) Polarimetric to Visible. (b) Polarimetric to Visible (DoG). (c) S0 to Visible. (d) S0 to Visible (DoG).

Table 2: The average SSIM results corresponding to different methods.

	Mahendran. *et al.* [15]	Riggan. *et al.* [20]	**GAN-VFS**
S0-Vis	0.2302	0.4342	**0.5491**
Poalr-Vis	0.2524	0.4423	**0.5603**
S0-Vis (DoG)	0.1785	0.2302	**0.2604**
Poalr-Vis (DoG)	0.1805	0.2404	**0.2672**

the gallery set and the reconstructed visible images from the corresponding polarimetric images in the same 30 subjects are regarded as the probe set. All the verification results are evaluated on the deep features extracted from the VGG-face model [17] without fine-tuning. Figure 6 plots the ROC curves corresponding to four experimental protocols. The area under the curve (AUC) and equal error rate (EER) results are reported in Table 3. From these results, it can be observed that the proposed method achieves the best performance evaluated on all four protocols. Meanwhile, it can be also observed that the reconstruction of the visible images from S0-Vis and Polar-Vis achieve better verifica-

tion results compared to their corresponding DoG versions. This is mainly due to the fact that while extracting VGG features from the DoG filtered images, the network was not fine-tuned based on DoG images. Therefore, features extracted for S0-Vis and Polar-Vis are better than features for their corresponding DoG versions.

Table 3: The average EER and AUC results corresponding to different methods.

		Raw	Mahendran *et al.* [15]	Riggan *et al.* [20]	**GAN-VFS**
AUC	S0-Vis	58.64%	59.25%	68.52%	**79.30%**
	Polar-Vis	50.35%	58.38%	75.83%	**79.90%**
	S0-Vis (DoG)	56.09%	57.04%	60.57%	**70.15%**
	Polar-Vis (DoG)	59.54%	59.02%	67.31	**70.41%**
EER	S0-Vis	43.96%	43.56%	34.36%	**27.34%**
	Polar-Vis	48.96%	44.56%	33.20%	**25.17%**
	S0-Vis (DoG)	46.48%	47.21%	41.27%	**36.25%**
	Polar-Vis (DoG)	43.33%	43.04%	37.23%	**34.58%**

5. Conclusion

We presented a new GAN-VFS network for synthesizing photo realistic visible face images from the corresponding polarimetric images. In contrast to the previous methods that regarded visible feature extraction and visible image reconstruction as two separate processes, we took a different approach in which these two steps are jointly optimized. Quantitative and qualitative experiments evaluated on a real polarimetric-visible dataset demonstrated that the proposed method is able to achieve significantly better results as compared to the recent state-of-the-art methods. In addition, an ablation study was performed to demonstrate the improvements obtained by different losses in the proposed method.

Acknowledgement

References

[1] T. Bourlai, N. Kalka, A. Ross, B. Cukic, and L. Hornak. Cross-spectral face verification in the short wave infrared (swir) band. In *ICPR*, pages 1343–1347. IEEE, 2010.

[2] R. Collobert, K. Kavukcuoglu, and C. Farabet. Torch7: A matlab-like environment for machine learning. In *BigLearn, NIPS Workshop*, 2011.

[3] E. Gonzalez-Sosa, R. Vera-Rodriguez, J. Fierrez, and V. M. Patel. Millimetre wave person recognition: hand-crafted vs. learned features. In *ISBA*, pages 1–7, 2017.

[4] I. Goodfellow, J. Pouget-Abadie, M. Mirza, B. Xu, D. Warde-Farley, S. Ozair, A. Courville, and Y. Bengio. Generative adversarial nets. In *NIPS*, pages 2672–2680, 2014.

[5] K. He, X. Zhang, S. Ren, and J. Sun. Delving deep into rectifiers: Surpassing human-level performance on imagenet classification. In *CVPR*, pages 1026–1034, 2015.

[6] S. Hu, N. J. Short, B. S. Riggan, C. Gordon, K. P. Gurton, M. Thielke, P. Gurram, and A. L. Chan. A polarimetric thermal database for face recognition research. In *CVPRW*, pages 119–126, 2016.

[7] S. Ioffe and C. Szegedy. Batch normalization: Accelerating deep network training by reducing internal covariate shift. In *ICML*, pages 448–456, 2015.

[8] P. Isola, J.-Y. Zhu, T. Zhou, and A. A. Efros. Image-to-image translation with conditional adversarial networks. *arxiv*, 2016.

[9] J. Johnson, A. Alahi, and L. Fei-Fei. Perceptual losses for real-time style transfer and super-resolution. In *ECCV*, pages 694–711. Springer, 2016.

[10] D. Kingma and J. Ba. Adam: A method for stochastic optimization. *arXiv preprint arXiv:1412.6980*, 2014.

[11] B. Klare and A. K. Jain. Heterogeneous face recognition: Matching nir to visible light images. In *ICPR*, pages 1513–1516, Aug 2010.

[12] C. Ledig, L. Theis, Huszár, et al. Photo-realistic single image super-resolution using a generative adversarial network. *arXiv preprint arXiv:1609.04802*, 2016.

[13] J. Lezama, Q. Qiu, and G. Sapiro. Not afraid of the dark: NIR-VIS face recognition via cross-spectral hallucination and low-rank embedding. *CoRR*, abs/1611.06638, 2016.

[14] C. Li and M. Wand. Precomputed real-time texture synthesis with markovian generative adversarial networks. In *ECCV*, pages 702–716, 2016.

[15] A. Mahendran and A. Vedaldi. Understanding deep image representations by inverting them. In *CVPR*, pages 5188–5196, 2015.

[16] F. Nicolo and N. A. Schmid. Long range cross-spectral face recognition: Matching swir against visible light images. *IEEE TIFS*, 7(6):1717–1726, Dec 2012.

[17] O. M. Parkhi, A. Vedaldi, and A. Zisserman. Deep face recognition. In *British Machine Vision Conference*, 2015.

[18] D. Pathak, P. Krahenbuhl, J. Donahue, T. Darrell, and A. A. Efros. Context encoders: Feature learning by inpainting. In *CVPR*, 2016.

[19] S. Reed, Z. Akata, X. Yan, L. Logeswaran, B. Schiele, and H. Lee. Generative adversarial text to image synthesis. *arXiv preprint arXiv:1605.05396*, 2016.

[20] B. S. Riggan, N. J. Short, S. Hu, and H. Kwon. Estimation of visible spectrum faces from polarimetric thermal faces. In *BTAS*, pages 1–7. IEEE, 2016.

[21] M. S. Sarfraz and R. Stiefelhagen. Deep perceptual mapping for thermal to visible face recognition. *arXiv preprint arXiv:1507.02879*, 2015.

[22] F. Schroff, D. Kalenichenko, and J. Philbin. Facenet: A unified embedding for face recognition and clustering. In *CVPR*, pages 815–823, 2015.

[23] N. Short, S. Hu, P. Gurram, K. Gurton, and A. Chan. Improving cross-modal face recognition using polarimetric imaging. *Optics letters*, 40(6):882–885, 2015.

[24] N. Short, S. Hu, P. Gurram, K. Gurton, and A. Chan. Improving cross-modal face recognition using polarimetric imaging. *Optics letters*, 40(6):882–885, 2015.

[25] Z. Wang, A. C. Bovik, H. R. Sheikh, and E. P. Simoncelli. Image quality assessment: from error visibility to structural similarity. *IEEE TIP*, 13(4):600–612, 2004.

[26] Y. Wen, K. Zhang, Z. Li, and Y. Qiao. A discriminative feature learning approach for deep face recognition. In *ECCV*, pages 499–515. Springer, 2016.

[27] S. Xie and Z. Tu. Holistically-nested edge detection. In *Proceedings of the IEEE International Conference on Computer Vision*, pages 1395–1403, 2015.

[28] H. Zhang and K. Dana. Multi-style generative network for real-time transfer. *arXiv preprint arXiv:1703.06953*, 2017.

[29] H. Zhang, V. Sindagi, and V. M. Patel. Image de-raining using a conditional generative adversarial network. *arXiv preprint arXiv:1701.05957*, 2017.