

Attribute-Guided Deep Polarimetric Thermal-to-visible Face Recognition

Seyed Mehdi Iranmanesh, Nasser M. Nasrabadi
West Virginia University

{seiranmanesh}@mix.wvu.edu, {nasser.nasrabadi}@mail.wvu.edu

Abstract

In this paper, we present an attribute-guided deep coupled learning framework to address the problem of matching polarimetric thermal face photos against a gallery of visible faces. The coupled framework contains two sub-networks, one dedicated to the visible spectrum and the second sub-network dedicated to the polarimetric thermal spectrum. Each sub-network is made of a generative adversarial network (GAN) architecture. We propose a novel Attribute-Guided Coupled Generative Adversarial Network (AGC-GAN) architecture which utilizes facial attributes to improve the thermal-to-visible face recognition performance. The proposed AGC-GAN exploits the facial attributes and leverages multiple loss functions in order to learn rich discriminative features in a common embedding subspace. To achieve a realistic photo reconstruction while preserving the discriminative information, we also add a perceptual loss term to the coupling loss function. An ablation study is performed to show the effectiveness of different loss functions for optimizing the proposed method. Moreover, the superiority of the model compared to the state-of-the-art models is demonstrated using polarimetric dataset.

1. Introduction

In recent years, there has been significant amount of research in Heterogeneous Face Recognition (HFR) [17]. The main issue in HFR is to match the visible face image to a face image that has been captured in another domain such as in the infrared spectrum [17], or the polarimetric thermal [7] domain. Infrared images are categorized into two major groups of reflection and emission. The reflection category, which contains near infrared (NIR) and shortwave infrared (SWIR) bands, is more informative about the facial details and it is very similar to the visible imagery. Due to this reflective phenomenology of the NIR and SWIR, there has been a significant performance on NIR-to-visible face

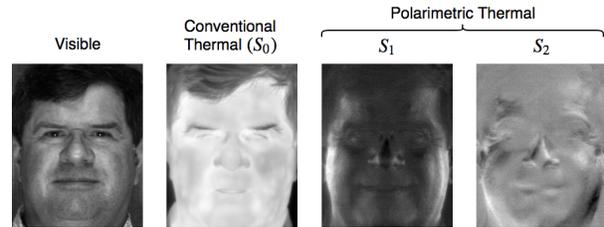


Figure 1. Visible spectrum and its corresponding conventional thermal (S_0), and polarimetric state information (S_1 and S_2) of a thermal image of a subject.

recognition accuracy [11] and to some extent for SWIR-to-visible face recognition accuracy [16].

The emission category contains the midwave infrared (MWIR) and longwave infrared (LWIR) bands and it is less informative [17] compared with the reflection category. Due to the significant phenomenological differences between the distribution of thermal and visible imagery, matching a thermal face against a gallery of visible faces becomes a challenging task. However, thermal-to-visible recognition is highly demanding because in the thermal data no active illumination is needed at night-time or low-light environments since the thermal imagery is based on the emission and originates from the underlying skin and depends on the individual's physiology. Many recent approaches on cross-modal problem have mainly focused on closing the gap between the two different modalities, but they have not explored the soft biometrics i.e., facial attributes.

Recently, via an emerging technology [7], the polarization state information of thermal emission has been exploited to provide additional geometrical and textural details, especially around the nose and the mouth, which complements the textural details of the conventional intensity-based thermal images. This additional information which is not available in the conventional intensity-based thermal imaging [7], is utilized in recent algorithms to enhance the cross thermal-to-visible face recognition [5]. Fig. 1 shows a visible image and its corresponding conventional thermal (S_0) and polarimetric state information (S_1 and S_2) images.

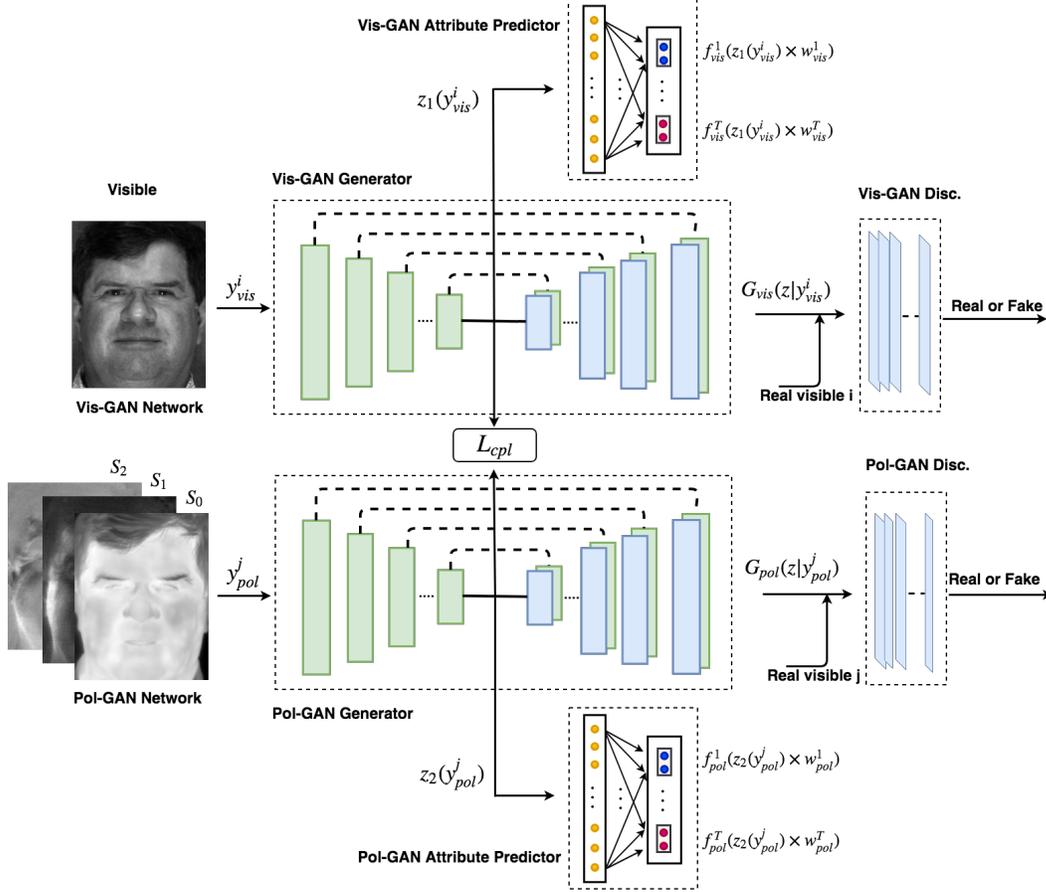


Figure 2. Proposed network using two GAN-based sub-networks (Vis-GAN and Pol-GAN) coupled by a contrastive loss function. Input to the Pol-GAN is the polarimetric data (S_0, S_1, S_2). The facial attributes are predicted from both sub-networks (Vis-GAN, Pol-GAN) in a multi-tasking paradigm.

There are some informative traits, such as age, gender, ethnicity, race, and hair color, which are not distinctive enough for the sake of recognition, but still can act as complementary information to other primary information, such as face and fingerprint. These traits, which are known as soft biometrics, can improve recognition algorithms while they are much cheaper and faster to acquire. They can be directly used in a unimodal system for some applications. Soft biometric traits have been utilized jointly with hard biometrics (face photo) for different tasks such as person identification or face recognition [3]. However, the soft biometric traits are considered to be available both during the training and testing phases. Our approach looks at this problem in a different way. We consider the case when soft biometric information does not exist during the testing phase, and our method can predict them directly in a *multi-tasking* paradigm.

Multi-task learning (MTL) and deep learning techniques has been vastly applied in computer vision and biometrics problems [13, 14, 15]. MTL basically attempts to solve cor-

related tasks concurrently with the help of knowledge sharing between tasks. In [1] MTL is employed to predict attributes such as age, gender, race, etc. Face photo can be viewed as having some positive or negative hidden relationship with some of its soft facial biometric traits.

In this paper, we propose an Attribute-Guided Coupled Generative Adversarial Network (AGC-GAN), which considers the Convolutional Neural Network (CNN) weight sharing followed by the dedicated weights which are responsible for learning the representative features for each specific face attribute. The network also tries to find the common embedding space between the polarimetric thermal and visible faces utilizing a coupling structure and an adversarial training. Optimizing the coupled network with the help of the facial attributes leads to a more discriminative embedding subspace and can be utilized to enhance the performance of the main task which is the heterogeneous face recognition.

To summarize, the following are our main contributions:

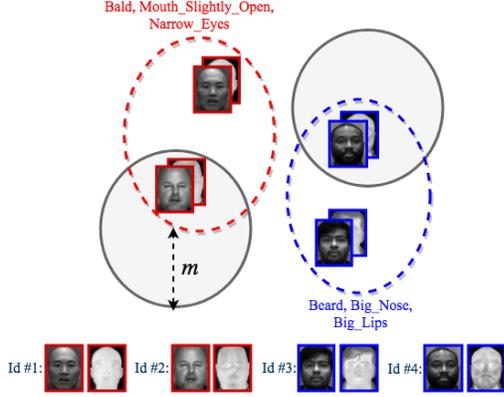


Figure 3. Visualization of the common latent subspace by leveraging facial attributes prediction loss function. Solid circles represent the contrastive margin and the dashed circles depict the attributes classification. For the sake of clarity the contrastive margin is depicted for two *Ids* out of four *Ids*.

- A novel polarimetric thermal-to-visible face recognition system is proposed based on an attribute-guided coupled GAN to synthesize visible faces from the polarimetric thermal images.
- A multi-tasking framework is proposed to predict facial attributes from the polarimetric thermal faces. To the best of our knowledge, no such demonstration has been proposed in the literature.
- Extensive experiments are conducted on ARL polarimetric facial database [7] and the proposed method is compared to recent state-of-the-art methods.

2. Preliminaries

2.1. Conditional Generative Adversarial Networks

The generative adversarial network consists of two networks, namely a generator and discriminator which compete with each other in a minimax game [4]. Conditional adversarial networks is an extension of generative adversarial networks in which both the generator and discriminator are conditioned on some auxiliary information y . The extra information y can be any kind of information such as class label or data from other modalities. The objective of the conditional GAN is the same as the classical GAN as follows:

$$\min_G \max_D E_{x \sim P_{data(x)}} [\log D(x|y)] + E_z \sim P_z [\log(1 - D(G(z|y)))] \quad (1)$$

3. Proposed Method

The proposed AGC-GAN is illustrated in Fig. 2. The proposed approach consists of two coupled generators and

two discriminators. Each generator is also responsible to predict facial attributes in a multi-tasking paradigm. In the following subsections we explain these modules in detail.

3.1. Deep Coupled Framework

The final objective of the proposed model is the identification of the polarimetric test faces against a gallery of visible faces which we do not have access to them during the training phase. For this reason, we couple two U-net networks [21] one is dedicated to the visible spectrum (Vis-GAN) and the other network is dedicated to the polarimetric spectrum (Pol-GAN). Each network performs a non-linear transformation of the input space. The final objective of our proposed AGC-GAN is to find the global deep latent features representing the relationship between polarimetric face images and their corresponding visible ones. In order to find a common latent embedding subspace between these two different domains we couple two networks (Vis-GAN and Pol-GAN) via a contrastive loss function [2]. This loss function (ℓ_{cont}) pulls the genuine pairs (i.e., a visible face image with its own corresponding polarimetric face images) towards each other in a common latent feature subspace and push the impostor pairs (i.e., a visible face image of a subject with another subject's polarimetric face images) apart from each other (see Fig. 2). Similar to [2], our contrastive loss is of the form:

$$\ell_{cont}(z_1(y_{vis}^i), z_2(y_{pol}^j), y_{cont}) = (1 - y_{cont})L_{gen}(D(z_1(y_{vis}^i), z_2(y_{pol}^j))) + y_{cont}L_{imp}(D(z_1(y_{vis}^i), z_2(y_{pol}^j))), \quad (2)$$

where y_{vis}^i is the input for the Vis-GAN (i.e., visible face image), and y_{pol}^j is the input for the pol-GAN (i.e., 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 $D(z_1(y_{vis}^i), z_2(y_{pol}^j))$ indicates the Euclidean distance between the embedded data in the common latent feature subspace. $z_1(\cdot)$ and $z_2(\cdot)$ are the deep CNN-based embedding functions, which transform y_{vis}^i and y_{pol}^j into a common latent embedding subspace. The binary label, y_{cont} , is assigned a value of 0 when both modalities, i.e., visible and polarimetric, form a genuine pair, or, equivalently, the inputs are from the same class ($cl^i = cl^j$). On the contrary, when the inputs are from different classes, which means they form an impostor pair, y_{cont} is equal to 1. In addition, L_{gen} and L_{imp} are defined as follows:

$$L_{gen}(D(z_1(y_{vis}^i), z_2(y_{pol}^j))) = \frac{1}{2} \|z_1(y_{vis}^i) - z_2(y_{pol}^j)\|_2^2 \quad \text{for } cl^i = cl^j, \quad (3)$$

and

$$L_{imp}(D(z_1(y_{vis}^i), z_2(y_{pol}^j))) = \frac{1}{2} \max(0, m - \|z_1(y_{vis}^i) - z_2(y_{pol}^j)\|_2^2) \text{ for } cl^i \neq cl^j, \quad (4)$$

where m is the contrastive margin. Therefore, the coupling loss function can be written as:

$$L_{cpl} = 1/N^2 \sum_{i=1}^N \sum_{j=1}^N \ell_{cont}(z_1(y_{vis}^i), z_2(y_{pol}^j), y_{cont}), \quad (5)$$

where N is the number of training samples. It should be noted that the contrastive loss function (5) considers the subjects' labels implicitly. Therefore, it has the ability to find a discriminative embedding space by employing the data labels in contrast to some other metrics such as the Euclidean distance. This discriminative embedding subspace would be useful in identifying a polarimetric probe photo against a gallery of visible face photos.

3.2. Multi-Attribute Prediction and Identification Task:

The objective of this model is to predict a set of attributes using a visible or polarimetric face images. Therefore, in this architecture a visible face image (polarimetric face images) is presented to the network as an input and a set of attributes are predicted. Suppose the input is a visible image $y_{vis}^i \in Y$, and its class label is $cl^i \in CL$ for $i = 1, \dots, N$ where N is the number of the training samples. Let the soft biometric traits, contain T different facial attributes or binary class labels. Therefore, in this framework we denote them as cl^t for $t = 1, \dots, T$. Learning multiple CNNs separately is not optimal since different tasks may have some hidden relationships with each other and may share some common features. This is supported by [24] where they train a CNN features for the face recognition task and they use it directly for the facial attribute prediction. Therefore, our network shares a big portion of its parameters among different attribute prediction tasks in order to enhance the performance of the recognition task. Thus, the loss function is as follows:

$$L_{avis} = 1/N \sum_{i=1}^N \sum_{t=1}^T \ell(f_{vis}^t(z_1(y_{vis}^i) \times w_{vis}^t), cl^{i,t}), \quad (6)$$

where ℓ is a proper loss function (e.g., cross entropy) and $f_{vis}^t(\cdot)$ is a binary classifier for the attribute t operated on the bottleneck of Vis-GAN (see Fig. 2). w_{vis}^t represents the remaining parameters which are learned separately for each facial attribute task.

The same procedure is performed in the other network (Pol-GAN) with polarimetric thermal images as input. The

Pol-GAN network is also responsible to estimate a set of facial attributes. Therefore, the loss function is:

$$L_{apol} = 1/N \sum_{j=1}^N \sum_{t=1}^T \ell(f_{pol}^t(z_2(y_{pol}^j) \times w_{pol}^t), cl^{j,t}), \quad (7)$$

where ℓ is the cross entropy loss function and $f_{pol}^t(\cdot)$ is a binary classifier for the attribute t operated on the bottleneck of Pol-GAN (see Fig. 2). w_{pol}^t represents the remaining features which are assigned separately for each facial attribute prediction task. The total attribute prediction loss function is:

$$L_a = L_{avis} + L_{apol}. \quad (8)$$

3.3. Generative Adversarial Loss

Let G_{vis} and G_{pol} denote the generators that synthesize the corresponding visible image from the visible and its polarimetric images, respectively. To synthesize the output and to make sure that the synthesized images generated by the two generators are indistinguishable from the corresponding ground truth visible image, we utilized the GAN loss function in [4]. As it is shown in Fig. 2, the first generator G_{vis} is responsible to generate a visible image when the network is conditioned on a visible image. On the other hand, the second generator G_{pol} tries to generate the visible image from the polarimetric images which is a more challenging task compared to the first generator. Therefore, the total loss for the coupled GAN is as follows:

$$L_{GAN} = L_{vis} + L_{pol}, \quad (9)$$

where the GAN loss function for the Vis-GAN sub-network is given as:

$$L_{vis} = \min_{G_{vis}} \max_{D_{vis}} E_{x^i \sim P_{vis}(x)} [\log D(x^i | y_{vis}^i)] + E_z \sim P_z [\log(1 - D(G(z | y_{vis}^i)))] \quad (10)$$

where y_{vis}^i is the visible image used as condition for the Vis-GAN and x^i is the real visible data. It should be noted that for the Vis-GAN the real data x^i and the condition y_{vis}^i are the same. Similarly the loss for the Pol-GAN is given as:

$$L_{pol} = \min_{G_{pol}} \max_{D_{pol}} E_{x^j \sim P_{vis}(x)} [\log D(x^j | y_{pol}^j)] + E_z \sim P_z [\log(1 - D(G(z | y_{pol}^j)))] \quad (11)$$

where y_{pol}^j is the polarimetric images used as condition for the Pol-GAN and x^j is the real visible data. It should be noted that x^i is the same as x^j if they refer to the same person ($cl^i = cl^j$) and otherwise they are not the same.

3.4. Overall Loss Function

The proposed approach contains the following loss functions: the Euclidean $L_{E_{vis}}$ and $L_{E_{pol}}$ losses which are enforced on the recovered visible images from the Vis-GAN and Pol-GAN sub-networks, respectively, are defined as follows:

$$L_{E_{vis}} = \|G_{vis}(z|y_{vis}^i) - x^i\|_2^2, \quad (12)$$

$$L_{E_{pol}} = \|G_{pol}(z|y_{pol}^j) - x^j\|_2^2, \quad (13)$$

$$L_E = L_{E_{vis}} + L_{E_{pol}}. \quad (14)$$

The L_{GAN} (9) loss is also added to generate sharper images. In addition, based on the success of the perceptual loss in low-level vision tasks [9], a perceptual loss is added to the Pol-GAN sub-network to generate a more realistic face photos as follows:

$$L_{P_{pol}} = \frac{1}{C_p W_p H_p} \sum_{c=1}^{C_p} \sum_{w=1}^{W_p} \sum_{h=1}^{H_p} \|V(G_{pol}(z|y_{pol}^j))^{c,w,h} - V(x^j)^{c,w,h}\|, \quad (15)$$

where x^j is the ground truth visible image, $G_{pol}(z|y_{pol}^j)$ is the output of Pol-GAN generator. $V(\cdot)$ represents a non-linear CNN transformation and C_p, W_p, H_p are the dimension of a particular layer in V . It should be noted that the perceptual loss is just used in the Pol-GAN sub-network. Similarly, we utilized a perceptual attribute loss which measures the difference between the facial attributes of the synthesized images and the real image. To extract attributes from a given visible face, we fine-tune the pre-trained VGG-Face [18] on ten annotated facial attributes as tabulated in Table 2. This network (attribute predictor) is trained separately from AGC-GAN. Afterward, this attribute predictor is utilized for perceptual attribute loss on Vis-GAN and Pol-GAN as follows:

$$L_{pavis} = \|A(G_{vis}(z|y_{vis}^i)) - A(x^i)\|_2^2, \quad (16)$$

$$L_{papol} = \|A(G_{pol}(z|y_{pol}^j)) - A(x^j)\|_2^2, \quad (17)$$

$$L_{pa} = L_{pavis} + L_{papol}, \quad (18)$$

where A is the fine-tuned VGG-Face attribute predictor network. L_{pa} is the total perceptual attribute loss function which composed of the perceptual attribute losses for the Vis-GAN (L_{pavis}) and Pol-GAN (L_{papol}) sub-networks.

Finally, the coupling loss function (5) is added to train both networks Vis-GAN and Pol-GAN jointly to make the embedding subspace of the mentioned networks as close as possible and to preserve a more discriminative and distinguishable shared space. Therefore, the total loss function is as follows:

$$L_T = L_{cpl} + \lambda_1 L_E + \lambda_2 L_{GAN} + \lambda_3 L_a + \lambda_4 L_{P_{pol}} + \lambda_5 L_{pa}, \quad (19)$$

where $\lambda_1, \lambda_2, \lambda_3, \lambda_4,$ and λ_5 are the hyper-parameters which weight different loss terms in the total loss function.

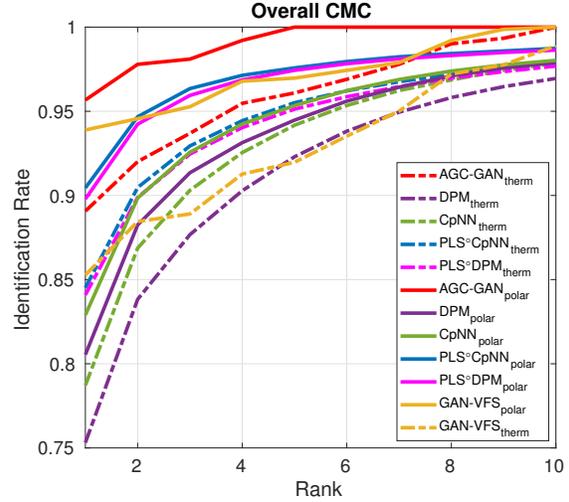


Figure 4. Overall CMC curves from testing AGC-GAN versus other baselines using polarimetric and thermal probe samples.

3.5. Testing Phase

During the testing phase, only the Pol-GAN is used. For a given test probe y_{pol}^t , Pol-GAN sub-network of the proposed AGC-GAN is employed to synthesize the visible image $G_{pol}(z|y_{pol}^t) = \hat{x}_{vis}^t$. Eventually, the identification of face recognition is done, by calculating the minimum Euclidean distance between the synthesized image from the polarimetric probe against the visible gallery images as follows:

$$x_{vis}^{t*} = \operatorname{argmin}_{x_{vis}^t} \|x_{vis}^t, \hat{x}_{vis}^t\|, \quad (20)$$

where \hat{x}_{vis}^t is the synthesized probe face image and x_{vis}^{t*} is the selected matching visible face image within the gallery of face images. In addition, the Pol-GAN sub-network can be employed to predict the facial attributes from the polarimetric face probe. The predicted facial attributes can also be used to narrow the search or identify a person of interest in a visible gallery of faces.

3.6. Implementation Details

A U-net structure [21] is employed as the network for the generator since it is able to address the vanishing gradient problem as well as capturing large receptive field. Also, a patch-based discriminator [8] is used in the proposed method and it is trained iteratively with the generator. The entire network is trained in Pytorch. For the sake of training AGC-GAN, the hyper-parameters for all the loss functions are set to one except for the perceptual loss $L_{p_{pol}}$ and perceptual attribute loss L_{pa} which is set to 0.5. For training we used Adam optimizer [10] with the first-order

momentum of 0.5, the learning rate of 0.0002, and batch size of 4. For the generator the ReLU activation, and for the discriminator the Leaky ReLU activation with the slope of 0.2 is considered. The perceptual loss is assessed on the relu3-1 layer of the pre-trained VGG model [12]. In order to fine-tune the attribute predictor network utilized for perceptual attribute loss, we manually annotated images with the attributes tabulated in Table 2.

4. Experiments

We evaluate the proposed face recognition method by comparing with several recent works [6, 22, 19, 7, 20, 23] on the ARL Multi-modal Face database [7].

Polarimetric Thermal Face dataset [7] comprises polarimetric thermal face images and their corresponding visible spectrum faces related to 60 subjects. Data was collected at three different distances: Range 1 (2.5 m), Range 2 (5 m), and Range 3 (7.5 m). At each range two different conditions, including baseline and expression are considered.

To increase the correlation between the two modalities of visible and thermal, each modality was preprocessed. We applied a band-pass filter so called difference of Gaussians (DoG), to emphasize the edges in addition to removing high and low frequency noise.

We pass S_0 , S_1 , and S_2 to the Pol-GAN’s three channels as the input as shown in Fig. 2. To train the network, the genuine and impostor pairs are constructed. The genuine pair is constructed for the same subject photos from the two different modalities. For the impostor pair, a different subject is selected for each modality. In general, the number of the generated impostor pairs are significantly larger than the genuine pairs. For the sake of balancing the training set, we consider the same number of genuine and impostor pair. After training the model, during the testing phase, only the polarimetric sub-network is used for the evaluation. For a given probe, the Pol-GAN sub-network is used to synthesize the visible image. Afterwards, the Euclidean distance is used to match the synthesize image to its closest image from the gallery. The ratio of the number of correctly classified subjects and the entire number of subjects is computed as the identification rate.

In each experiment the dataset is partitioned to the training and testing sets randomly. The same set of training and testing are used to evaluate PLS [6], DPM [22], CpNN [19], PLS◦DPM [7], PLS◦CpNN [20], GAN-VFS [23], and the proposed AGC-GAN network. Fig. 4 shows the overall cumulative matching characteristics (CMC) curves for our proposed method and the other state-of-the-art methods over all the three different ranges as well as the expressions data at Range 1. For the sake of comparison, in addition to the polarimetric thermal-to-visible face recognition performance, Fig. 4 also shows the results for the conventional

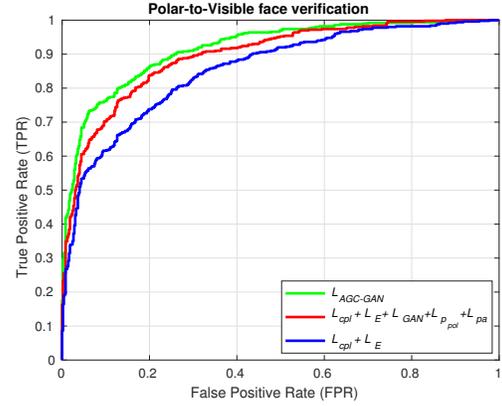


Figure 5. The ROC curves corresponding to the ablation study.

thermal-to-visible face recognition for some of the methods, namely PLS◦DPM, PLS◦CpNN, CpNN, and AGC-GAN. In the conventional thermal-to-visible face recognition, all the mentioned methods exactly follow the same procedure as before, with only using S_0 modality. Fig. 4 illustrates that exploiting the polarization information of the thermal spectrum enhances cross-spectrum face recognition performance compared to the conventional one. Fig. 4 also shows the superior performance of our approach compared to the state-of-the-art methods. In addition, our method could achieve perfect accuracy (100%) at Rank-4 and above.

Table 1 tabulates the Rank-1 identification rates for five different scenarios: overall (which corresponds to Fig. 4), Range 1 expressions, Range 1 baseline, Range 2 baseline, and Range 3 baseline. In our proposed approach, exploiting polarization information enhance the Rank-1 identification rate by 2.24%, 5.16%, 4.65%, and 9.44% for Range 1 baseline, Range 1 expression, Range 2 baseline, and Range 3 baseline compared to the conventional thermal-to-visible face recognition. This table reveals that using deep coupled generative adversarial network technique with the contrastive loss function as well as utilizing facial attributes to transform different modalities into a distinctive common embedding subspace is superior to the other embedding techniques such as PLS◦CpNN. It also shows the effectiveness of our method in exploiting polarization information to improve the cross-spectrum face recognition problem.

5. Ablation Study

In order to illustrate the effect of adding different loss functions and their improvement in our proposed framework, we perform a study with the following evaluations using the polarimetric dataset: 1) Polar-to-visible using the coupled framework with using only $L_{cpl} + L_E$ losses, 2) Polar-to-visible using the proposed framework with $L_{cpl} + L_E + L_{GAN} + L_{p_{pol}} + L_{p_a}$ loss functions, and 3) Polar-

Table 1. Rank-1 identification rate for cross-spectrum face recognition using polarimetric thermal and conventional thermal (S_0) probe imagery.

| Scenario | Rank-1 Identification Rate | | | | | | | |
|------------------|----------------------------|--------|--------|--------|---------|----------|---------|---------------|
| | Probe | PLS | DPM | CpNN | PLS◦DPM | PLS◦CpNN | GAN-VFS | AGC-GAN |
| Overall | Polar | 0.5867 | 0.8054 | 0.8290 | 0.8979 | 0.9045 | 0.9382 | 0.9654 |
| | Therm | 0.5305 | 0.7531 | 0.7872 | 0.8409 | 0.8452 | 0.8561 | 0.8925 |
| Expressions | Polar | 0.5658 | 0.8324 | 0.8597 | 0.9565 | 0.9559 | 0.9473 | 0.9733 |
| | Therm | 0.6276 | 0.7887 | 0.8213 | 0.8898 | 0.8907 | 0.8934 | 0.9217 |
| Range 1 Baseline | Polar | 0.7410 | 0.9092 | 0.9207 | 0.9646 | 0.9646 | 0.9653 | 0.9883 |
| | Therm | 0.6211 | 0.8778 | 0.9102 | 0.9417 | 0.9388 | 0.9412 | 0.9659 |
| Range 2 Baseline | Polar | 0.5570 | 0.8229 | 0.8489 | 0.9105 | 0.9187 | 0.9263 | 0.9643 |
| | Therm | 0.5197 | 0.7532 | 0.7904 | 0.8578 | 0.8586 | 0.8701 | 0.9178 |
| Range 3 Baseline | Polar | 0.3396 | 0.6033 | 0.6253 | 0.6445 | 0.6739 | 0.8491 | 0.9068 |
| | Therm | 0.3448 | 0.5219 | 0.5588 | 0.5768 | 0.6014 | 0.7559 | 0.8124 |

to-visible with all the loss functions in the proposed framework (19).

We plot the receiver operation characteristic (ROC) curves corresponding to the mentioned three different settings of the framework in the task of face verification. As it is shown in Fig. 5 the L_{GAN} has an important rule in the enhancement of our proposed approach by transforming the polarimetric modality to the visible one. Moreover, adding facial attribute prediction loss enhances the face recognition performance. The reason behind this is because using facial attributes loss in addition to contrastive loss function leads to a more discriminative embedding space and this leads to a better face recognition performance. Consider a polarimetric subject with $Id\#2$ (see Fig. 3). The contrastive loss function causes the corresponding visible images from $Id\#2$ to move closer to $Id\#2$'s polarimetric and other Ids' visible images to move farther away. Now, using the contrastive loss function in conjunction with the attribute classification loss function makes $Id\#1$ to move closer to $Id\#2$ since they share the same set of attributes (see Fig. 3). In other words, it differentiates between different impostors of $Id\#2$. The same procedure is performed for the other identities during the training process. Fig. 3 visualizes the overall concept of adding facial attributes prediction loss function. As it is depicted, addition of attribute prediction loss leads to a more discriminative embedding subspace. This leads to a better face recognition performance as it is shown by the ROC curves in Fig. 5.

6. Attribute Prediction From Polarimetric Thermal

One of the benefits of the proposed AGC-GAN is predicting facial attributes directly from polarimetric thermal modality. These attributes can be utilized directly or can be fused with other modalities to enhance recognition performance. In order to illustrate the effectiveness of the proposed method we performed attribute prediction in four different scenarios: 1) Attribute prediction of visible images

with the VGG-Face based attribute predictor (Network A in Sec. 3.4). 2) Attribute prediction of polarimetric images with the attribute predictor A . 3) Attribute prediction of polarimetric images with the fine-tuned attribute predictor A . In this case, we fine-tuned the attribute predictor A with the annotated polarimetric images and used it for the task of attribute prediction in the testing phase. 4) Attribute prediction of the polarimetric images using the Pol-GAN attribute predictor from the proposed AGC-GAN (see Fig. 2). Table 2 shows the result of the prediction for the four mentioned frameworks. Although, fine-tuning the attribute predictor A increased the prediction performance (framework #3), but still its performance is less than our proposed framework. The proposed framework could outperform the other methods in predicting polarimetric face attributes and it has a comparable performance with the accuracy of predicted attributes from the visible face images (framework #1).

7. Conclusion

We have introduced a novel approach to exploit facial attributes information for the purpose of polarimetric thermal-to-visible face recognition. The AGC-GAN contains two GAN-based sub-networks dedicated to visible and polarimetric input images. The proposed network is capable of transforming the visible and polarimetric thermal modalities into a common discriminative embedding subspace and synthesizing the visible photos from the embedding subspace. We simultaneously minimize the cost functions due to the facial attribute identification in addition to other cost functions in order to increase the inter-personal variations between different subjects with different sets of facial attributes in the latent feature subspace. This leads to a more discriminative embedding subspace. An ablation study was performed to demonstrate the enhancement obtained by different losses in the proposed method. We compared our method with state-of-the-art polarimetric thermal-to-visible face recognition methods and showed the superiority of our proposed method over them.

Table 2. Attribute prediction of the polarimetric face images using the proposed method and other frameworks and comparing it with the attribute prediction of visible faces.

| Facial Attributes | Arched_Eyebrows | Big_Lips | Big_Nose | Bushy_Eyebrows | Bald | Mustache | Narrow_Eyes | Beard | Mouth_Slightly_Open | Young |
|----------------------------|-----------------|----------|----------|----------------|------|----------|-------------|-------|---------------------|-------|
| Visible Input (Net A) | 96.7 | 98.4 | 99.1 | 95.9 | 99.3 | 99.4 | 95.7 | 98.9 | 97.7 | 96.9 |
| Polar Input (Net A) | 53.7 | 55.8 | 57.1 | 51.2 | 58.9 | 62.8 | 54.4 | 59.8 | 57.6 | 52.7 |
| Polar Input (fine-tuned A) | 78.2 | 83.9 | 85.3 | 80.7 | 88.4 | 89.3 | 79.3 | 88.9 | 81.9 | 76.5 |
| Polar Input (Pol-GAN) | 89.6 | 95.3 | 96.6 | 90.4 | 96.2 | 95.2 | 91.9 | 94.8 | 93.7 | 91.1 |

References

- [1] A. H. Abdunabi, G. Wang, J. Lu, and K. Jia. Multi-task CNN model for attribute prediction. *IEEE Transactions on Multimedia*, 17(11):1949–1959, 2015.
- [2] S. Chopra, R. Hadsell, and Y. 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*, volume 1, pages 539–546. IEEE, 2005.
- [3] A. Dantcheva, P. Elia, and A. Ross. What else does your biometric data reveal? a survey on soft biometrics. *IEEE Transactions on Information Forensics and Security*, 11(3):441–467, 2016.
- [4] I. Goodfellow, J. Pouget-Abadie, M. Mirza, B. Xu, D. Warde-Farley, S. Ozair, A. Courville, and Y. Bengio. Generative adversarial nets. In *Advances in neural information processing systems*, pages 2672–2680, 2014.
- [5] K. P. Gurton, A. J. Yuffa, and G. W. Videen. Enhanced facial recognition for thermal imagery using polarimetric imaging. *Optics letters*, 39(13):3857–3859, 2014.
- [6] S. Hu, J. Choi, A. L. Chan, and W. R. Schwartz. Thermal-to-visible face recognition using partial least squares. *JOSA A*, 32(3):431–442, 2015.
- [7] 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 *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops*, pages 119–126, 2016.
- [8] P. Isola, J.-Y. Zhu, T. Zhou, and A. A. Efros. Image-to-image translation with conditional adversarial networks. *CVPR*, 2017.
- [9] J. Johnson, A. Alahi, and L. Fei-Fei. Perceptual losses for real-time style transfer and super-resolution. In *European Conference on Computer Vision*, pages 694–711. Springer, 2016.
- [10] D. P. 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 *Pattern Recognition (ICPR), 2010 20th International Conference on*, pages 1513–1516. IEEE, 2010.
- [12] A. Krizhevsky, I. Sutskever, and G. E. Hinton. Imagenet classification with deep convolutional neural networks. In *Advances in neural information processing systems*, pages 1097–1105, 2012.
- [13] R. Mehrizi, X. Peng, D. N. Metaxas, X. Xu, S. Zhang, and K. Li. Predicting 3-d lower back joint load in lifting: A deep pose estimation approach. *IEEE Transactions on Human-Machine Systems*, 49(1):85–94, 2019.
- [14] R. Mehrizi, X. Peng, Z. Tang, X. Xu, D. Metaxas, and K. Li. Toward marker-free 3d pose estimation in lifting: A deep multi-view solution. In *2018 13th IEEE International Conference on Automatic Face & Gesture Recognition (FG 2018)*, pages 485–491. IEEE, 2018.
- [15] R. Mehrizi, X. Peng, X. Xu, S. Zhang, D. Metaxas, and K. Li. A computer vision based method for 3d posture estimation of symmetrical lifting. *Journal of biomechanics*, 69:40–46, 2018.
- [16] F. Nicolo and N. A. Schmid. Long range cross-spectral face recognition: matching SWIR against visible light images. *IEEE Transactions on Information Forensics and Security*, 7(6):1717–1726, 2012.
- [17] S. Ouyang, T. Hospedales, Y.-Z. Song, and X. Li. A survey on heterogeneous face recognition: Sketch, infra-red, 3d and low-resolution. *arXiv preprint arXiv:1409.5114*, 2014.
- [18] O. M. Parkhi, A. Vedaldi, A. Zisserman, et al. Deep face recognition. In *BMVC*, volume 1, page 6, 2015.
- [19] B. S. Riggan, C. Reale, and N. M. Nasrabadi. Coupled auto-associative neural networks for heterogeneous face recognition. *IEEE Access*, 3:1620–1632, 2015.
- [20] B. S. Riggan, N. J. Short, and S. 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*, pages 1–7. IEEE, 2016.
- [21] O. Ronneberger, P. Fischer, and T. Brox. U-net: Convolutional networks for biomedical image segmentation. In *International Conference on Medical image computing and computer-assisted intervention*, pages 234–241. Springer, 2015.
- [22] M. S. Sarfraz and R. Stiefelhagen. Deep perceptual mapping for thermal to visible face recognition. *arXiv preprint arXiv:1507.02879*, 2015.
- [23] H. Zhang, V. M. Patel, B. S. Riggan, and S. Hu. Generative adversarial network-based synthesis of visible faces from polarimetric thermal faces. In *Biometrics (IJCB), 2017 IEEE International Joint Conference on*, pages 100–107. IEEE, 2017.
- [24] Y. Zhong, J. Sullivan, and H. Li. Face attribute prediction using off-the-shelf CNN features. In *Biometrics (ICB), 2016 International Conference on*, pages 1–7. IEEE, 2016.