Cross-spectrum thermal to visible face recognition based on cascaded image synthesis

Khawla Mallat[§], Naser Damer[⋆†], Fadi Boutros[⋆†], Arjan Kuijper[⋆†], Jean-Luc Dugelay[§]
[§]Digital Security Department, EURECOM, Sophia Antipolis, France
[⋆]Fraunhofer Institute for Computer Graphics Research IGD, Darmstadt, Germany
[†]Mathematical and Applied Visual Computing, TU Darmstadt, Darmstadt, Germany
Email: khawla.mallat@eurecom.fr

Abstract

Face synthesis from thermal to visible spectrum is fundamental to perform cross-spectrum face recognition as it simplifies its integration in existing commercial face recognition systems and enables manual face verification. In this paper, a new solution based on cascaded refinement networks is proposed. This method generates visible-like colored images of high visual quality without requiring large amounts of training data. By employing a contextual loss function during training, the proposed network is inherently scale and rotation invariant. We discuss the visual perception of the generated visible-like faces in comparison with recent works. We also provide an objective evaluation in terms of cross-spectrum face recognition, where the generated faces were compared against a gallery in visible spectrum using two state-of-the-art deep learning based face recognition systems. When compared to the recently published TV-GAN solution, the performance of the face recognition systems, OpenFace and LightCNN, was improved by a 42.48% (i.e. from 10.76% to 15.37%) and a 71.43% (i.e. from 33.606% to 57.612%), respectively.

1. Introduction

Predominantly, law enforcement and security systems have been focused in the visible spectrum. This pertains to a large number of applications from biometrics, access control systems to video surveillance. Particularly, face recognition, one of the most important tasks in these aforementioned applications, has achieved remarkable performances due to the uprise of deep learning and the abundant amount of available data. However, face recognition systems are prone to fail when employed in unconstrained conditions. Among the main challenges in visible spectrum-based security systems, variable or low illumination conditions have proven to be some of the major weaknesses of such systems, particularly that most of the security breaches occur during night time. A promising approach for detecting potential threats in total darkness is using thermal imagery. Thermal imagery detects electromagnetic radiations in the medium wave MWIR (3 - 8μm) and long wave infrared spectrum LWIR (8 - 15μm) in which most of the heat energy is emitted by any object [17]. Therefore, it is possible to acquire a crisp image without any external source of illumination, based on subtle differences in temperature.

Thermal imaging technology has drastically advanced during the last couple of decades, while thermal cameras have evolved to become affordable and user friendly. Even though thermal imaging solutions are significantly advancing, they still suffer from poor performances due to low image resolution, lack of color, and poor texture and geometric information. Inasmuch as exploring thermal imagery is considerably new, only a few public databases are available, thus it cannot profit from deep learning technologies to develop reliable face recognition systems operating in thermal spectrum. Particular studies of face recognition have thoroughly focused on bringing visible and thermal spectra together to benefit from the advantages of each. This discipline, referred to as cross-spectrum face recognition, aims in our case of study to identify a person imaged in thermal spectrum from a gallery containing face images acquired in the visible spectrum.

In this work, we focus on image synthesis strategy for cross-spectrum face recognition, consisting in generating visible-like images from thermal captures that will be matched against a gallery of visible faces. Opting for this strategy is essential to enable its integration in existing face recognition systems as well as manual face verification by human examiners. We propose using cascaded refinement networks coupled with contextual loss to synthesize high quality colored visible images from thermal acquisitions.

978-1-7281-3640-0/19/$31.00 ©2019 IEEE

The proposed solution is computationally-efficient and inherently scale and rotation invariant, thus it does not require large aligned training sets. Using the VIS-TH database [13] of simultaneous face acquisition in both visible and thermal spectra, we validate our face synthesis system in different poses and occlusion scenarios. An evaluation of the generated faces in cross-spectrum face recognition application is performed using two different state-of-the-art systems OpenFace [2] and LightCNN [23].

2. Related work

First attempts to investigate face synthesis from thermal to visible were conducted by Li et al. [12]. Their work presents a learning-based framework that takes advantage of the local linearity in the spatial domain of the image as well as in the image manifolds. Then, they apply Markov random field to organize the image patches and improve the estimated visible-like face images. Dou et al. [7] used Canonical Correlation Analysis (CCA) to extract the features in order to find one-to-one mapping between thermal and visible faces. The relationship of the two feature spaces is then learnt using Locally Linear Regression. Finally, Locally Linear Embedding is utilized to reconstruct the visible-like face from the converted thermal features.

In the wake of the recent advances in deep learning, several works were based on Generative Adversarial Neural network (GAN) to synthesize visible not only from thermal inputs [26, 22], but also from near-infrared [21, 10], and polarimetric data [28, 27]. GANs, first introduced by I. Goodefellow in [9], can learn to generate from any distribution of data through a contest of two neural networks: generator and discriminator. The generator aims to maximize the probability of making the discriminator classify its output as real. While the discriminator pushes the generator to generate more realistic data.

Different models can also be used for similar conversion. For examples, deep convolutional Generative Adversarial network (DCGAN)[16] and Boundary Equilibrium Generative Adversarial Networks (BEGAN) [3]. DCGAN introduced the Convolution Neural Network (CNN) into the discriminator and the generator. BEGAN introduced equilibrium factor that controls the model training by balancing the discriminator and generator. These GAN models significantly improved the training stability, but they did not improve the generated images quality. However, some GAN approaches such as Cycle-Consistent Adversarial Networks (CycleGAN) [29] and Image-to-Image Translation with Conditional Adversarial Nets (Pix2Pix) [11] were able to achieve higher resolution images, but it ends with adding more complexity to the model. CycleGAN consists of four neural networks (two generators and two discriminators). Training such a big model is computationally costly and requires large databases, that are unavailable for an application like the one dealt with in this paper, to achieve satisfactory results.

Zhang et al. [28] considered synthesizing colored faces from thermal images with various head poses and occlusion with eyeglasses. This work used Conditional GANs inspired from pix2pix system [11] coupled with a closed-set face recognition loss that led to preserve the face identity information. A cross-spectrum face recognition evaluation is performed, using the pre-trained MatConvNet VGG-based model, and reported a performance improvement of 14.88% compared to pix2pix system. A recent work by Wang et al. [22] derived from CycleGAN model [29] incorporates facial landmark detector loss that depicts face identity preserving features. This system was evaluated using FaceNet pretrained on public available visible datasets, and has improved cross-spectrum face recognition performance by 3% compared to CycleGAN system. However, this work is different from our framework since its aim is to generate visible face images in gray scale and discarded face generation under challenging conditions such as head pose and occlusion.

3. Proposed method

To generate images, we propose to base our approach on the cascaded refinement network (CRN) [4]. We chose the CRN as the basic block for our image generation as it considers multi-scale information and based on training a limited number of parameters. This allows for a higher resolution generation and less data size dependency in comparison to solutions based on GAN. CRN is a convolutional neural network that consists of inter-connected refinement modules. Each module consists of only three layers, input, intermediate, and output layer. The first module considers the lowest resolution space (4x4 in our case). This resolution is duplicated in the successor modules until the last module (128x128 in our case), matching the target image resolution. For more detailed description of the CRN, one can refer to [4]. An illustration of the image synthesis approach using CRN is shown in Figure 1. The input thermal images are processed at different scales and fed into the next level in the cascade along with the thermal image at the next scale. Finally, the targeted image (visible in this case) is synthesized.

To control the training of our CRN network, we used the contextual loss function (CL) [14]. This choice is based on our need for: a) a loss function that is robust to not well aligned data (as in our use-case where input face images are not uniformly aligned), and b) neglect outliers on the pixel level (in comparison to pixel level loss [11, 24]). Gramm loss [8] can satisfy the two aforementioned conditions, however, unlike CL, it does not constrain the content of the generated image as it describes the image globally.

The CL function can be calculated between the source

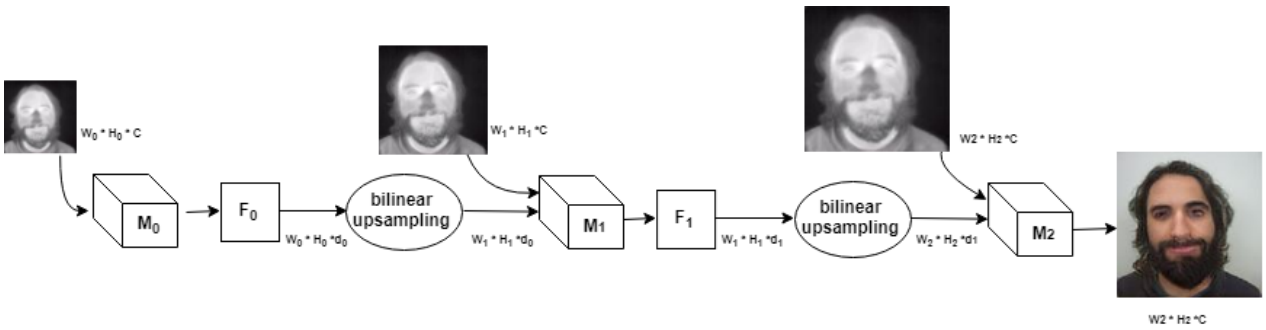


Figure 1: The CRN-based multi-scale approach to transform the thermal image into a visible-like image. Here only three consecutive modules are shown as an example. These cascaded modules can be repeated until reaching the targeted destination resolution.

(thermal image) and the generated images, or between the target (visible ground-truth image) and the generated images. The source-generated loss aims at saving the details of the source image such as detailed boundaries. The target-generated loss maintains the properties of the target image in the generated image, e.g. target image style. In our case, as will be presented in the next section, the source and target training image pairs are of identical faces. Therefore, the target-generated loss also maintains the detailed properties of the object in the source image.

Both losses were calculated between image embeddings extracted by a pre-trained VGG19 [20] network trained on the ImageNet database [6]. The total loss is calculated as given in [14] and formulated as:

$$\begin{aligned} L_{CX}(s,t,g,l_s,l_t) =& \lambda_1(-log(CX(\Phi_1^{l_s}(g),\Phi^{l_s}(s)))) + \\ & \lambda_2(-log(CX(\Phi_2^{l_t}(g),\Phi^{l_t}(t)))), \end{aligned} \quad (1)$$

where s, t, and g are the are the source, target, and generated images respectively. CX is the rotation and scale invariant contextual similarity [14]. Φ is a perceptual network, VGG19 in our work. $\Phi^{l_s}(x)$, $\Phi^{l_t}(x)$ are the embeddings vectors extracted from the image x at layer l_s and l_t of the perceptual network respectively. Here l_s is the `conv4_2` and l_t is the `conv3_2` and `conv4_2` layers, as motivated in [14]. In our implementation, $\lambda_1 = 0.01$ and $\lambda_2 = 0.99$ by checking the resulting generated image visually. Moreover, as mentioned earlier, because the pairs of source and target images are of identical faces, and thus the loss weighted by λ_2 maintains the structural details of the source image. The training was run for 40 epochs, batch size of one, and 1e-4 learning rate.

4. Experiments and results

In this section, we describe the database used for the development and the evaluation of our proposed solution. Then, we detail the evaluation protocol used to assess the generated data in cross-spectrum face recognition task. Finally, we present the baselines followed by an analysis of the obtained results.

4.1. Database

We used the VIS-TH face database [13] for the development and the evaluation of our solution. The database is publicly available and contains face images in both visible spectrum with pixel resolution of 1920×1080 and thermal spectrum of pixel resolution 160×120 with a spectral response range of 7.5 - 13.5μm. Unlike the few existing databases of visible and thermal face, this database is acquired simultaneously using the dual sensor camera Flir DUO R [1] considering a wide range of facial variations. The database contains in total 2100 images collected from 50 subjects of different ages, gender, and ethnicities. For the evaluation, we have considered 5 subsets of the database split per facial variation as follow:

- **Neutral:** One single capture acquired with neutral expression, frontal pose and standard illumination.
- **Expression:** 6 captures acquired with different face expressions: smiling, angry, sad, surprised, blinking, yawning.
- **Head pose:** 4 captures acquired with different head poses: up, down, right at 30°, left at 30°.
- **Occlusion:** 5 captures acquired with varying occlusions: eyeglasses, sunglasses, cap, mouth occluded by hand, eye occluded by hand.
- **Illumination:** 5 captures acquired with different illuminations: room light, rim light, key light, fill light, all lights on, all lights off.

4.2. Evaluation protocol

Images, from both visible and thermal spectrum, were normalized and sampled to 128×128. Enabling an evaluation of our solution in hands-on scenarios, and considering that face alignment in thermal spectrum still remains a challenge itself, the face images were not aligned, thus they contained slight variable shifts.

Face images from 45 subjects, except for the ones acquired in total darkness, were used for training the face synthesis network. The thermal face images from the remaining 5 subjects were fed to the trained model to synthesis the visible-like images. This experiment was performed 10 times in order to synthesis all the images contained in the database without overlapping the test and train images or identities.

For evaluating the synthesized faces when used in cross-spectrum face recognition task, we measured the recognition accuracy of two selected state-of-the-art face recognition systems:

OpenFace [2] is an implementation of face recognition system using deep neural networks based on Google's FaceNet [19] architecture. The OpenFace network is trained using the combination of the two largest public face databases CASIA-WebFace [25] and FaceScrub [15]. The evaluation of OpenFace model provided competitive performances compared to private state-of-the-art systems. We use the OpenFace pretrained model to map faces into 128-dimension embeddings. Then, nearest neighbors algorithm is performed using Euclidean distance to discriminate matching samples.

LightCNN [23] is a new implementation of CNN for face recognition designed to have fewer trainable parameters and to handle noisy labels. This network introduces a new concept of maxout activation in each convolutional layer, called Max-Feature-Map, for feature filter selection. This network has achieved better performance than CNNs while reducing computational costs and storage space. When evaluated on the LFW database, LightCNN achieved face recognition accuracy of 99.33% outperforming OpenFace that obtained 92.92% of accuracy. We used the learned network with 29-layers to obtain embeddings of 256-dimension from face images. Embeddings extracted from gallery and probe templates are compared using cosine similarity.

4.3. Baselines

The performance of our image synthesis solution in cross-spectrum face recognition is compared to the following baselines:

Visible We perform face recognition in the visible spectrum, by considering the neutral face image as gallery and the rest of the facial variations as probe images. This will report the performance of the face recognition systems used in this paper for the evaluation of the generated images. Besides, this baseline will depict the utility of thermal to visible face synthesis in hands-on scenarios, in particular when the face is acquired in poorly lit environment.

Thermal Here, we conduct cross-spectrum face recognition without any modifications applied to the thermal data. Simply put, we consider as gallery set the neutral face image acquired in visible spectrum and as probe set all the other face variations in acquired in thermal spectrum. This baseline will quantify the gap between the two spectra.

Isola et al. [11], referred to as Pix2Pix, learns the mapping from one domain to another, by training a Conditional GAN using Least Absolute Deviations (L1) loss function. The generator is based on U-Net [18] architecture, an encoder-decoder with skip connections between mirrored layers in the encoder and decoder stacks. Whilst the discriminator aims to classify real images from generated ones. The training was run for 85 epochs, batch size of one, and 2e-4 learning rate.

Zhang et al. [26], have designed a network, called TV-GAN, notably to generate visible-like face images from thermal captures. This work is inspired from Pix2Pix, as it uses the same exact network for the generator. However, the authors proposed a multi-task discriminator, that doesn't only classify real from generated images, but also performs a closed-set face recognition to obtain identity loss. This aims to generate visible-like images while preserving identity information from the thermal inputs. The training was run for 65 epochs, batch size of one, and 2e-4 learning rate.

4.4. Results

The images in Fig. 2 illustrate, in each row, a sample from different facial variations of synthesized visible-like images from thermal captures. The first column shows the source thermal faces. From the second to the fourth column, we present visible-like faces synthesized using pix2pix model by Isola et al. [11], TV-GAN model by Zhang et al. [26] and finally our model based on cascaded refinement network, respectively. A detailed analyses of the generated image quality is presented in [5]. The last column shows the ground truth visible faces.

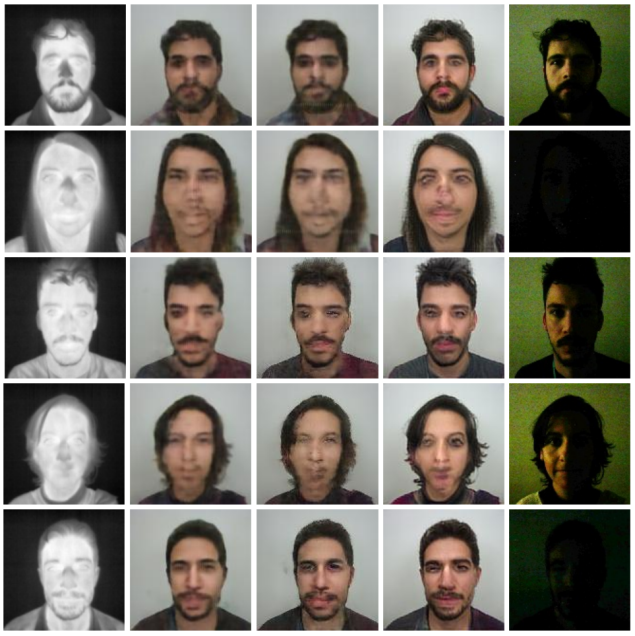
The different face images with frontal face pose were synthesized with satisfying visual quality. Although we note that our proposed model has succeeded in generating more informative details (e.g. eyes, mouth) compared to the pix2pix and TV-GAN results, it does not always generate the correct attributes such as race and gender. We can observe that all generated visible-like faces differ in skin color from the ground-truth images, and this applies to all synthesis models. This is due to the fact that thermal images do not contain texture and color information, thus, it is difficult to infer the skin color tone from the thermal prints. Another visual distortion can be noted on the visible-like samples gen-

erated by our proposed model in the second and the fourth row of Figure 2. These samples show some added facial-hair around the mouth and the jaws area. This observation can be reasoned by the unbalanced distribution of gender representation within the training data. Third and sixth row display samples from different head poses, where we can remark major artifacts in the synthesized visible-like faces when compared to the frontal head pose. As for images acquired with occlusion, illustrated in the fourth and seventh row, they were synthesized in relative good quality. However, we perceive some confusion in generating faces with eyeglasses. This can be justified by the fact that the training data contains samples with eyeglasses and others with sunglasses that both have similar thermal print. Overall, it is noteworthy that our proposed model provides visible-like faces that are the most visually pleasing compared to pix2pix and TV-GAN models.

Thermal Isola et al.[11] Zhang et al.[26] Ours Ground truth

Figure 2: Selected samples of synthesized face images under challenging scenarios

In order to evaluate the generated visible-like face images, we have performed cross-spectrum face recognition



Thermal Isola et al.[11] Zhang et al.[26] Ours Ground truth

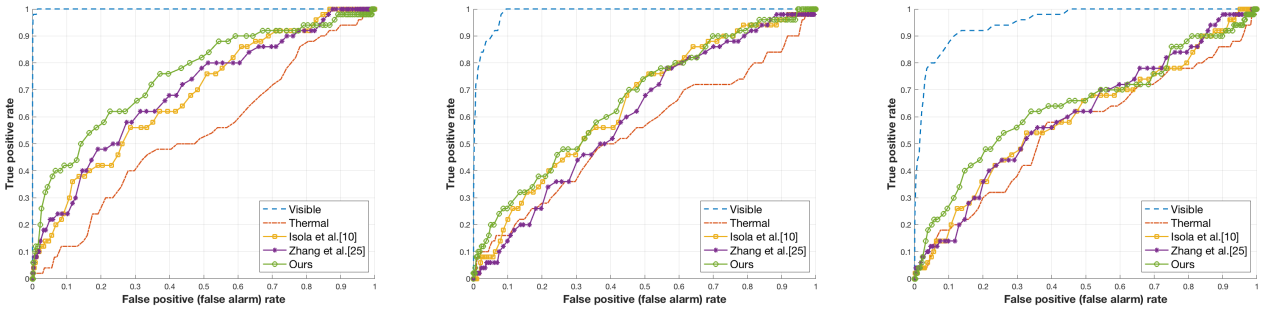
Figure 3: Samples of generated images acquired in total darkness

using two different systems. The evaluation experiment consists in comparing, in the first place, the generated neutral face against the ground truth and then matching the generated faces from each of the facial variation subsets against the visible neutral face. We report, in table 1 and table 2, the recognition accuracy of the OpenFace and LightCNN, respectively. To get a deeper understanding of the performance of the two face recongition systems used to evaluate the results obtained, we plot the receiver operating characteristic (ROC) curves, in Fig. 4 and Fig. 5, corresponding to some selected samples from different face variations.

We note from the reported results that all synthesis models outperformed the thermal, which proves the efficiency of synthesizing visible-like in reducing spectral gap between visible and thermal data. TV-GAN reports better performances that pix2pix confirming the efficacy of the identity loss in preserving the subject identity when generating visible-like images. Foremost, our proposed solution, based on cascaded refinement networks, outperforms all the models by a large margin, particularly observed on LightCNN results, and that applies to all facial variations. This is mainly due to the limitations of GAN networks that are known for being data hungry. However, our system succeeded in generating relatively high quality visible-like images despite the limited size of the training data. Furthermore, both pix2pix and TV-GAN models are based on L1 loss function making them very sensitive to image misalign-

	Visible	Thermal	Isola et al. [11]	Zhang et al. [26]	Ours
Neutral	100	4	8	20	**20**
Expression	97.66	3.33	7.66	11	**17.33**
Head Pose	75.5	2.5	4	8	**9.5**
Occlusion	80	2	7.2	8.4	**10**
Illumination	80.8	3.2	10.4	11.6	**20**
Average	86.79	3.01	8.49	10.76	**15.37**

Table 1: Cross-spectrum face recognition accuracy across multiple facial variations using OpenFace system



(a) (b) (c)

Figure 4: ROC curves of cross-spectrum face recognition based on OpenFace system for selected samples from: (a) expression variation, (b) head pose variation, (c) occlusion variation.

ment. Whilst our proposed system uses contextual loss which makes it inherently scale and rotation invariant.

To highlight the main motivation of this work, we display, in Fig. 3, samples that are acquired in operative scenarios of thermal sensors usage, where face images were captured in total darkness. As we were expecting, the poor or absent illumination does not impact the generated visible-like images. In fact, we succeeded in synthesizing images with informative facial attributes that are absent in the visible spectrum.

Table 3 reports the accuracy of OpenFace and LightCNN face recognition systems. We plot also, in Fig. 6 the ROC curves of the two evaluation systems in the absolute dark condition. We can clearly observe that our proposed model not only outperforms other face synthesis models but also it provides significantly higher performance compared to the visible spectrum. This affirms the efficacy of face synthesis from thermal to visible in most of the challenging scenarios such as poorly lit environments.

5. Conclusion

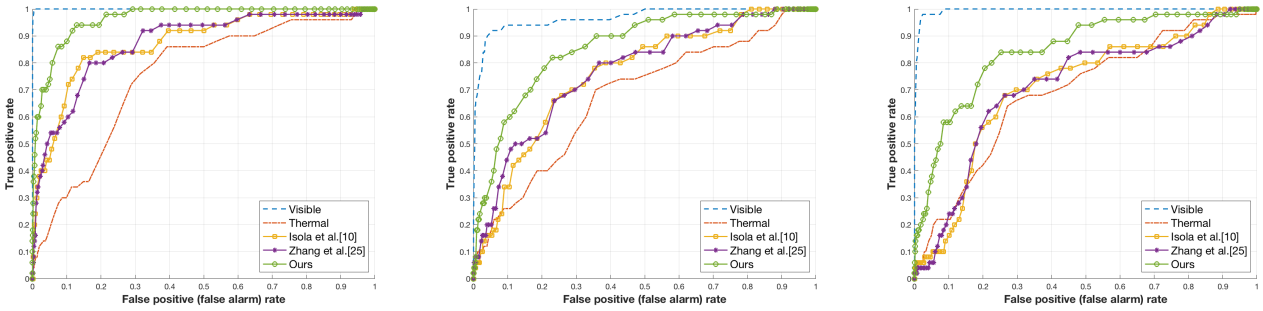
Although several efforts have been devoted lately for face synthesis from thermal to visible spectrum, it is still challenging considering the shortage of the available data designed for this task. We present, in this paper, a novel solution based on cascaded refinement networks, that succeeded in generating high-quality color visible image, trained on limited size database. The proposed network is based on the use of contextual loss function, enabling it to be inherently scale and rotation invariant. Despite the existence of challenging facial variations such as occlusions, expression, head pose and illumination, our solution has produced the most visually pleasing synthesized face images when compared to existing work. We also performed applicability evaluation of our solution in cross-spectrum face recognition task. The reported results have shown that our system outperforms recent face synthesis systems. Underlining the motivation of face synthesis from thermal to visible spectrum, we have proved that face recognition performance reported on the synthesized images is significantly higher than the one reported on visible spectrum when operated in poorly lit environments, as it was improved by 37.5% (i.e. from 16% to 22%) and 33.33% (i.e. from 42% to 56%) evaluated by OpenFace and LightCNN, respectively.

6. Acknowledgement

EURECOM's Research activities in dual visible and thermal imagery are partly supported through funding from FR FUI COOPOL and the European Union's Horizon 2020 within the PROTECT project.This work was also supported by the German Federal Ministry of Education and Research (BMBF) as well as by the Hessen State Ministry for Higher Education, Research and the Arts (HMWK) within CRISP.

	Visible	Thermal	Isola et al. [11]	Zhang et al. [26]	Ours
Neutral	100	32	48	54	**82**
Expression	99.66	23	37.33	38.33	**67.66**
Head Pose	80.5	12.5	14.5	15.5	**30**
Occlusion	98.8	14.4	16.4	25	**44.8**
Illumination	87.2	15.6	29.6	35.2	**63.6**
Average	95.232	19.5	29.166	33.606	**57.612**

Table 2: Cross-spectrum face recognition accuracy across multiple facial variations using LightCNN system

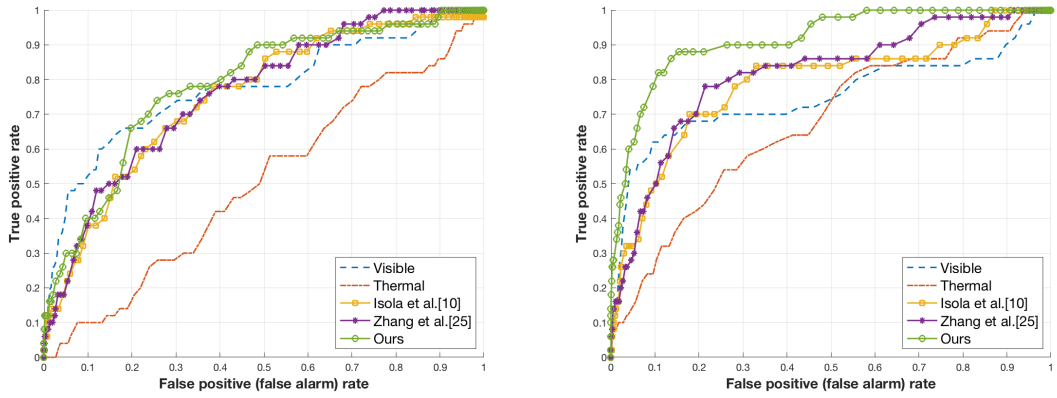


(a) (b) (c)

Figure 5: ROC curves of cross-spectrum face recognition based on LightCNN system for selected samples from: (a) expression variation, (b) head pose variation, (c) occlusion variation.

	Visible	Thermal	Isola et al. [11]	Zhang et al. [26]	Ours
OpenFace	16	2	10	14	**22**
LightCNN	42	16	28	36	**56**

Table 3: Cross-spectrum face recognition accuracy in operative scenario where samples were acquired in total darkness



(a) (b)

Figure 6: ROC curves of cross-spectrum face recognition in dark environment: (a) OpenFace system (b) LightCNN system.

References

[1] Flir systems. http://www.flir.com/.

[2] B. Amos, B. Ludwiczuk, and M. Satyanarayanan. Openface: A general-purpose face recognition library with mobile applications. Technical report, CMU-CS-16-118, CMU School of Computer Science, 2016.

[3] D. Berthelot, T. Schumm, and L. Metz. Began: boundary equilibrium generative adversarial networks. *European Conference of Computer Vision Workshops (ECCVW)*, 2018.

[4] Q. Chen and V. Koltun. Photographic image synthesis with cascaded refinement networks. In *IEEE International Conference on Computer Vision, ICCV 2017, Venice, Italy, October 22-29*, pages 1520–1529. IEEE Computer Society, 2017.

[5] N. Damer, F. Boutros, K. Mallat, F. Kirchbuchner, J.-L. Dugelay, and A. Kuijper. Cascaded generation of high-

quality color visible face images from thermal captures. In *27th European Signal Processing Conference, EUSIPCO 2019, A Coruña, Spain, 2019*. IEEE (under review), 2019.

[6] J. Deng, W. Dong, R. Socher, L.-J. Li, K. Li, and L. Fei-Fei. ImageNet: A Large-Scale Hierarchical Image Database. In *CVPR09*, 2009.

[7] M. Dou, C. Zhang, P. Hao, and J. Li. Converting thermal infrared face images into normal gray-level images. pages 722–732. Asian Conference on Computer Vision, 2007.

[8] L. A. Gatys, A. S. Ecker, and M. Bethge. Image style transfer using convolutional neural networks. In *2016 IEEE Conference on Computer Vision and Pattern Recognition, CVPR 2016, Las Vegas, NV, USA, June 27-30, 2016*, pages 2414–2423. IEEE Computer Society, 2016.

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

[10] A. C. Guei and M. A. Akhloufi. Deep generative adversarial networks for infrared image enhancement. *Proc. SPIE*, 10661:10661 – 10661 – 12, 2018.

[11] P. Isola, J. Zhu, T. Zhou, and A. A. Efros. Image-to-image translation with conditional adversarial networks. In *2017 IEEE Conference on Computer Vision and Pattern Recognition, CVPR 2017, Honolulu, HI, USA, July 21-26, 2017*, pages 5967–5976. IEEE Computer Society, 2017.

[12] J. Li, P. Hao, C. Zhang, and M. Dou. Hallucinating faces from thermal infrared images. In *Proceedings of the International Conference on Image Processing, ICIP 2008, October 12-15, 2008, San Diego, California, USA*, pages 465–468. IEEE, 2008.

[13] K. Mallat and J. Dugelay. A benchmark database of visible and thermal paired face images across multiple variations. In *2018 International Conference of the Biometrics Special Interest Group (BIOSIG)*, pages 1–5, Sep. 2018.

[14] R. Mechrez, I. Talmi, and L. Zelnik-Manor. The contextual loss for image transformation with non-aligned data. In *Computer Vision - ECCV 2018 - 15th European Conference, Munich, Germany, September 8-14, 2018, Proceedings*, volume 11218 of *Lecture Notes in Computer Science*, pages 800–815. Springer, 2018.

[15] H. Ng and S. Winkler. A data-driven approach to cleaning large face datasets. In *2014 IEEE International Conference on Image Processing, ICIP 2014, Paris, France, October 27-30, 2014*, pages 343–347, 2014.

[16] A. Radford, L. Metz, and S. Chintala. Unsupervised representation learning with deep convolutional generative adversarial networks. *arXiv preprint arXiv:1511.06434*, 2015.

[17] M. Rai, T. Maity, and R. Yadav. Thermal imaging system and its real time applications: a survey. *Journal of Engineering Technology*, pages 290–303, 2017.

[18] O. Ronneberger, P. Fischer, and T. Brox. U-net: Convolutional networks for biomedical image segmentation. In N. Navab, J. Hornegger, W. M. Wells, and A. F. Frangi, editors, *Medical Image Computing and Computer-Assisted Intervention – MICCAI 2015*, pages 234–241, Cham, 2015. Springer International Publishing.

[19] F. Schroff, D. Kalenichenko, and J. Philbin. Facenet: A unified embedding for face recognition and clustering. In *IEEE Conference on Computer Vision and Pattern Recognition, CVPR 2015, Boston, MA, USA, June 7-12, 2015*, pages 815–823, 2015.

[20] K. Simonyan and A. Zisserman. Very deep convolutional networks for large-scale image recognition. *CoRR*, abs/1409.1556, 2014.

[21] L. Song, M. Zhang, X. Wu, and R. He. Adversarial discriminative heterogeneous face recognition. AAAI Conference on Artificial Intelligence, 2018.

[22] Z. Wang, Z. Chen, and F. Wu. Thermal to visible facial image translation using generative adversarial networks. *IEEE Signal Processing Letters*, 25(8):1161–1165, Aug 2018.

[23] X. Wu, R. He, Z. Sun, and T. Tan. A light cnn for deep face representation with noisy labels. *IEEE Transactions on Information Forensics and Security*, 13(11):2884–2896, 2018.

[24] L. Xu, J. S. J. Ren, C. Liu, and J. Jia. Deep convolutional neural network for image deconvolution. In *Advances in Neural Information Processing Systems 27: Annual Conference on Neural Information Processing Systems 2014, December 8-13 2014, Montreal, Quebec, Canada*, pages 1790–1798, 2014.

[25] D. Yi, Z. Lei, S. Liao, and S. Z. Li. Learning face representation from scratch. *CoRR*, abs/1411.7923, 2014.

[26] H. Zhang, V. M. Patel, B. S. Riggan, and S. Hu. Generative adversarial network-based synthesis of visible faces from polarimetric thermal faces. In *2017 IEEE International Joint Conference on Biometrics (IJCB)*, pages 100–107, Oct 2017.

[27] H. Zhang, B. S. Riggan, S. Hu, N. J. Short, and V. M. Patel. Synthesis of high-quality visible faces from polarimetric thermal faces using generative adversarial networks. *International Journal of Computer Vision*, 2018.

[28] T. Zhang, A. Wiliem, S. Yang, and B. Lovell. TV-GAN: Generative adversarial network based thermal to visible face recognition. In *2018 International Conference on Biometrics (ICB)*, pages 174–181, Feb 2018.

[29] J.-Y. Zhu, T. Park, P. Isola, and A. A. Efros. Unpaired image-to-image translation using cycle-consistent adversarial networkss. In *Computer Vision (ICCV), 2017 IEEE International Conference on*, 2017.