Thermal and Cross-spectral Palm Image Matching in the Visual Domain by Robust Image Transformation

Ewelina Bartuzi[1], Naser Damer[2,3]
[1]Research and Academic Computer Network (NASK), Warsaw, Poland
[2]Fraunhofer Institute for Computer Graphics Research IGD, Darmstadt, Germany
[3]Mathematical and Applied Visual Computing, TU Darmstadt, Darmstadt, Germany
Email: ewelina.bartuzi@nask.pl, naser.damer@igd.fraunhofer.de

Abstract

Synthesizing visual-like images from those captured in the thermal spectrum allows for direct cross-domain comparisons. Moreover, it enables thermal-to-thermal comparisons that take advantage of feature extraction methodologies developed for the visual domain. Hand based biometrics are socially accepted and can operate in a touchless mode. However, certain deployment scenarios requires captures in non-visual spectrums due to impractical illumination requirements. Generating visual-like palm images from thermal ones faces challenges related to the nature of hand biometrics. Such challenges are the dynamic nature of the hand and the difficulties in accurately aligning hand's scale and rotation, especially in the understudied thermal domain. Building such a synthetic solution is also challenged by the lack of large-scale databases that contain images collected in both spectra, as well as generating images of appropriate resolutions. Driven by these challenges, this paper presents a novel solution to transfer thermal palm images into high-quality visual-like images, regardless of the limited training data, or scale and rotational variations. We proved quality similarity and high correlation of the generated images to the original visual images. We used the synthesized images within verification approaches based on CNN and hand crafted-features. This allowed significantly improved the cross-spectral and thermal-to-thermal verification performances, reducing the EER from 37.12% to 16.25% and from 3.04% to 1.65%, respectively in both cases when using CNN-based features.

1. Introduction

Hand biometrics has been used since the XIX century, initially in the form of fingerprints [1], then geometric features [2, 3], and palmprints [4, 5, 6, 7]. Solutions based on vein patterns [8, 9, 10] are now widely used, especially in banking. Systems based on hand characteristics have many advantages, mainly related to the simple acquisition process and usability. Acquisition can be performed in a contactless manner, resulting in a hygienic and convenient process that can enable continuous verification [11]. The palm side of the hand is also rarely exposed in whole (less imitation attack-prone), and is socially acceptable. Recent works [12, 13] confirmed the discriminatory information content in the thermal hand images. More importantly, these thermal features are independent of external illumination variations and it is hard to reconstruct heat maps to perform presentation attacks on such a system [13]. On the other hand, thermal representations are highly dependent on temperature changes caused by variations in the body metabolism and other physiological processes. Such a challenge can be handled by developing a comparison pipeline that is robust to such variations, e.g. a thermal-to-visual image conversion approach that considers such variations.

To compare palm images captured in two different spectrum, one can either learn to compare them directly, or convert them (or one of them) to a common spectral representation. This is performed by learning to convert one of the images to the spectral representation of the other. Moreover, even comparing images captured with the same spectrum (e.g. thermal) can benefit from such conversion, by applying the comparison on generated images in a domain that have certain advantages. For example, converting images into the visual domain enables manual comparison, as well as taking advantage of feature extraction models trained on the largely available visual training images. Cross-spectral verification and spectral transformation of hand image were not addressed previously in the literature.

The recent image spectral transformation methods relaid on the recent advancements around the idea of generative adversarial networks (GAN). While the first GAN solutions lacked stability, more advanced designs, such as Boundary Equilibrium Generative Adversarial Networks (BEGAN)

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

and Deep Convolutional Generative Adversarial network (DCGAN), offered more stability but failed to produce images of high resolution. This issue was addressed by the Cycle-Consistent Adversarial Networks (CycleGAN) and Image-to-Image Translation with Conditional Adversarial Nets (Pix2Pix), which increased the complexity of the network and thus its training data needs. Requiring the conversion of thermal palm images to high resolution visual-like images in an alignment-free and scale-invariant manner without the needed for large training data, motivated the novel image transformation offered in this paper.

To address these challenges, we successfully utilized the Cascaded Refinement Network (CRN) to build our image transformation without the need of a large database. To target the alignment-free and scale-invariant needs, among others, we used the contextual loss function to learn an accurate and robust transformation of the content regardless of the mentioned factors. This resulted in a high quality visual-like image generation, even when the thermal information was weak, noisy, or blurry. We showed that the visual-like generated images had similar quality properties, as well as high correlation, to the targeted visual ones.

We successfully used the visual-like generated images along with handcrafted features and re-trained CNN features to perform thermal-to-thermal and thermal-to-visual comparisons, after transferring the thermal images into the visual-like domain. Both verification scenarios benefited from the proposed transformation by reducing the resulted EER in both cases by more than 45% in comparison to pre-transformation EER, when using CNN-based features.

2. Related work

Previous works are seen here under two categories. First is the hand image recognition in general, with a focus on the efforts made in thermal hand recognition. Second is the cross-spectral image transformation with a focus on its deployment in biometrics.

Hand biometrics: Identity recognition based on the hand is one of the longest used characteristics in biometrics. In the years 1892-1895, Galton developed biometrics based on fingerprints [1], currently used in a wide range of applications, from forensics to logical access control. Beyond fingerprints, hand images are collected using sensors sensitive to different ranges of light wavelength. The analysis of hand images acquired in visible light is, explicitly or implicitly, based on the main lines, minutiae, and skin texture patterns [14]. Some of the earlier feature extraction algorithms in this case are based on Principal Component Analysis (PCA) [15, 16], bank of Gabor filters [17], texture descriptors such as Local Binary Patterns (LBP)[17], Scale-invariant Feature Transform (SIFT) [18, 19, 5], Histogram of Oriented Gradients (HOG) [20], and Compcode [21, 22].

Hand thermal image biometrics is an understudied field, and was initially discussed in segmentation tasks. One of the first recognition approaches based on hand thermal maps was described by Czajka and Bulwan [23]. They used features extracted by PCA and Linear Discriminant Analysis and implemented a Support Vector Machine (SVM) based classifier (EER = 6.67%). Bartuzi *et al.* [12] used two texture descriptors, namely Binarized Statistical Image Features (BSIF) and LBP, and employed CNN-based approach. Obtained results have shown that thermal maps are characterized by poor stability. For intra-session comparisons it is possible to obtain 100% correct recognition results, however, the EER reached 17.17% when comparing images captured under different conditions.

There are also methods based on images collected in different ranges of wavelength. Here the goal is to increase the information content by combining information embedded in different types of photos, or allow cross-domain comparison between different sensors. Rowe *et al.* [24] and Samoil and Anushkevich [25, 26] fused results on the score-level, while Y. Hao *et al.* [27] proposed a method for data fusion at pixel-level. Faundez-Zanuy *et al.* [28] collected a multi-spectral database of hand images but reported only inter-spectral verification performance [28] and later proposed a thermal hand image segmentation approach [29]. Non of the previous works considered the cross-spectral comparison of hand images, nor the hand image transformation from one spectral domain to another. Both issues are the core contributions in this work.

Cross-spectral image transformation: Cross-spectral biometric comparison can be performed by cross-spectral matching, i.e. direct comparison of captures from different capture spectrum [30], or cross-spectral generation, i.e. converting one of the compared captures to the spectrum domain of the other capture. The availability of huge visual image databases allowed to train high-performing feature extraction networks optimized for the visual captures.

The study of transferring the image into a different spectral domain in biometric applications focused on face recognition. One of the first works in this direction was presented by Li *et al.* [31]. They proposed learning-based framework that exploits the local linearity in the spatial domain of the image and also in the image manifolds followed by applying Markov random field to improve the hallucinated visible face. More recently, many researchers exploited the development of the Generative Adversarial Neural networks (GAN). Riggan *et al.* [32] proposed a convolution neural network that aims to map edge-based features, locally and globally, from the polarimetric to the gray visible domain.

Synthesizing colored faces from thermal images was recently proposed by Zhang *et al.* [33] using Conditional Generative Adversarial Neural network. Given that GANs are data hungry, they had to use data augmentation to train the model, resulting in generated images with high level

of visible artifacts. Moreover, the work used aligned images and did not mention the generated images resolution . Different models can also be used for similar conversion. For examples, deep convolutional Generative Adversarial network (DCGAN)[34] and Boundary Equilibrium Generative Adversarial Networks (BEGAN) [35]. These GAN models significantly improved the training stability, but it did not improve the generated images quality. Where both models trained on (64×64 px) image size. However, some GAN approaches such as Cycle-Consistent Adversarial Networks (CycleGAN) [36] and Image-to-Image Translation with Conditional Adversarial Nets (Pix2Pix) [37] were able to achieve a higher resolutions images, but it ends with adding more complexity to the model. CycleGAN consists of four neural networks (two generators and tow 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.

3. Experimental data

In this work we use the *Tecnocampus Hand Image Database* (THID) [28, 29], which contains hand images obtained in three different wavelength ranges. In our experiments, two subsets of this database were used:

• **thermal images – oT**, acquired by thermal camera Testo 882-3 with sensor of range from 3 to $14\mu m$ (320×240 px),

• **visible light images – oV**, collected by digital camera ($380 - 750nm$, 640×480 px).

Data was collected from 111 subjects (100 subject in the first published database) in five sessions organized every two weeks except the last one, which took place after a month. Two thermal images and two visible light images were acquired per session for both side of right hands. We analyzed only palm side in our experiments. Figure 1, the first two columns present examples of images from *THID*.

4. Image transformation

This section presents one of the main contribution of this work, the alignment and scale invariant visual-like image generation from thermal images. In the following, we motivate and present our transformation methodology along with samples of the generated images. We also prove the quality similarity and high correlation between the generated images and the original visual ones.

Generation methodology: The novel solution we propose to generate visual-like palm images from thermal captures was motivated by the requirements and challenges associated with hand recognition. These can be listed as:
a) Hands have a high dynamic nature in comparison to other biometric characteristics. For example, the face may contain expression changes but has quite static boundaries. Therefore, the hand transformation method should focus on the entire image containing the object rather than the content of a defined, relatively stable, region as in face captures.
b) The hand scale within the image is variable and dependent on the capture setup. This can be solved by putting unrealistic constrains on the capture process, or automatic scale correction based on landmarks detection, which is a relatively understudied issue.
c) Aligning hand images is also a problematic issue because of the dynamic nature of the hand (mentioned earlier).
d) The generated images have to be of a realistic resolution. This limits the use of some GAN-based image generation solutions that suffer from resolution limitations.

The previous challenges are seen from the deployment perspective of a cross-spectral hand biometric. Other challenges can face the creation (training) of a cross-spectrum image generation. Some of these are:
e) The limited availability of perfectly aligned, simultaneously captured, multi-spectral palm images database is an issue facing training such a conversion, as the target image does not completely corresponds to the source image.
f) Training solutions for other biometric modalities, such as visual face, can depend on publicly available databases that contains thousands of different subjects. Due to the nature of hand biometrics (hard to be collected from search engines), the available databases are quite limited. This case is more dominant in cross-spectrum databases.

The previous points motivated the novel design of our transformation solution. To be able to generate high resolution images while having relatively small training data, we based our solution on the cascaded refinement network (CRN) [38]. The CRN considers multi-scale information and is based on training a relatively small number of parameters, leading to lower need of training data. The CRN is a convolutional neural network that consists of interconnected refinement modules. Each module consists of only three layers, input, intermediate, and output layer. The first module considers the lowest resolution space (4×4 in our case). This resolution is duplicated in the successor modules until the last module, matching the target image resolution. The second to last module expects two inputs, the output representation of the previous module and the input image at the specific resolution of the module. The target image (generated) resolution in out experiments was set to 512×512 pixels. The use of the CRN targeted the problem-specific constrains (d) and (f).

The CRN training is controlled by a loss function. The choice of this loss function targets our constrains (a, b, c, and e). The loss function should be invariant to exact scale, alignment, and exact shape of the hand (opened/closed fingers). This also inherently covers the training with not perfectly aligned pairs of visual and thermal images. This requires a loss function that neglects outliers on the pixel level

(in comparison to pixel-level loss [37, 39]). The Gramm loss [40] can satisfy the other requirements, however, it does not constrain the content of the generated image as it describes it globally, which is problematic for image generation. To achieve the required properties, we chose the contextual loss function (CL) [41].

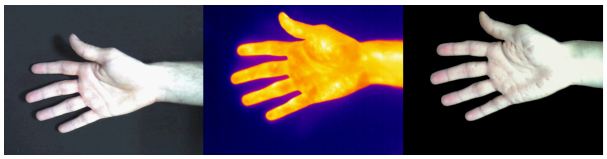
The CL function is calculated between the source and the generated images, or between the target (ground-truth) and the generated images. The source-generated loss aims at saving the details of the source image such as detailed boundaries. The target-generated loss maintains the properties of the target image in the generated image, e.g. target image style and content. In our case, the source (thermal) and target (visible-light) training image pairs are of identical (yet, not aligned or have the same scale) hand images.

Both losses were calculated between image embeddings extracted by a pre-trained VGG-19 [42] network trained on the ImageNet database [43]. The total loss is calculated as given in [41] and formulated as:

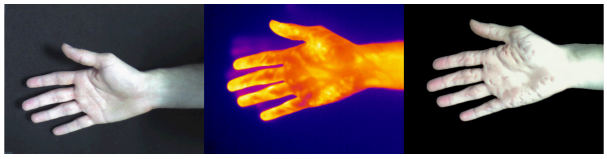
$$L_{CX}(s,t,g,l_s,l_t) = \lambda_1(-log(CX(\Phi_1^{l_s}(g),\Phi^{l_s}(s)))) + \lambda_2(-log(CX(\Phi_2^{l_t}(g),\Phi^{l_t}(t)))), \quad (1)$$

where s, t, and g are the are the source, target, and generated images respectively. CX is the rotation and scale invariant contextual similarity [41]. 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 [41] and utilized in a similar manner in face images in [44]. In our implementation, $\lambda_1 = 0.01$ and $\lambda_2 = 0.99$ by checking the resulting generated image visually. That was motivated to capture the style of the target image (visual), but maintain the main properties of the source image, shape and location.

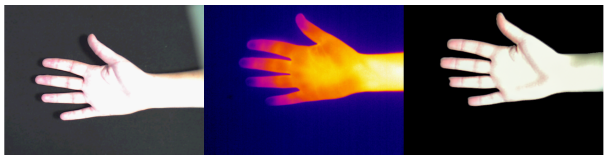
Generated dataset: Thermal and visual original images used for training are linearly interpolated to the size of 512×512 pixels to match our network input. No alignment or palm size related scaling are performed. All images were treated as three channel images by the network. The data (thermal and visual) is split into four parts based on the identity. The total of 111 identities produced 3 splits with 28 identities and one with 27 identities. For each training iteration, three of these parts were used to train our transformation model, then this model was used to transform the thermal images in the remaining data split into visual-like images. This was performed 4 times to test our generation on all the images in the database without overlapping the test and train images or identities. We trained the solution on a system with i5-6500 CPU 16 GB RAM and NVIDIA GTX 1050 Ti 4 GB. The training took 40 epochs, batch size of one, and 1e-4 learning rate. The training taken slightly under 6 hours. The generated images are linearly interpolated from the generation resolution (512×512 pixels), into the original aspect ratio (640×480 pixels). To remove any artifacts outside the hand area, a contour is used to detect the largest object (the hand) and mask the rest of the image. This resulted in a visual-like generated image for each of the thermal images in the database.



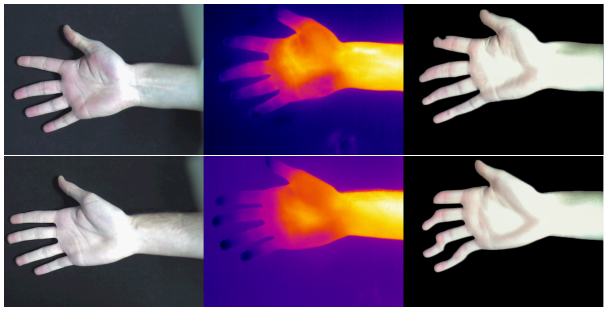
(a) '*Warm*' palm.



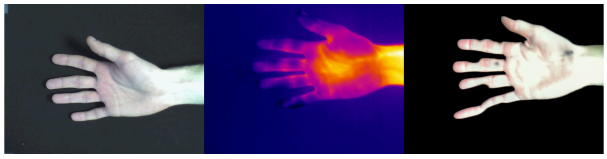
(b) Palm with distinct texture on thermal image.



(c) Palm with cold fingertips.



(d) Palm with complete cold fingers.



(e) Palm with large cold regions.

(f) Palm with heat-shade effect.

Figure 1: Example palm images categorized with characteristic features of thermal image. From left to right: original visible light images, original thermal images, and generated visible-like images from thermal spectrum.

Samples of the generated visual-like (gV) images are shown in Figure 1, along with the thermal images used for the generation and the associated visual image in the database. The visual similarity between the generated visual-like images and the corresponding original visual ones are noticeable. The transformation approach is able to deal successfully with variations of the thermal images, such as warmer than average thermal captures (1a), distinctive texture thermal images (1b), colder fingertips (1c), and heat-shade effect (1f). Even in extreme scenarios where the thermal signal is fading in larger areas of the hand (1e and 1d), generation results are realistic except in areas where the thermal information is completely missing.

Image quality: One of the success metrics of the image generation is to produce images that have similar quality properties to the target images. Here, the generated images (gV) are not a reproduction of the target images (oV), including desired scale and location differences influenced by the source image (thermal). This motivates reporting no-reference image quality metrics, rather than reference-related metrics. We report two unlearned metrics and two learned metrics. The unlearned ones are the Global Contrast Factor (GCF) [45] and the sharpness [46], both previously suggested for face capture quality [47]. The considered learned quality metrics are the Natural Image Quality Evaluator (NIQE) [48] and the Blind/Referenceless Image Spatial Quality Evaluator (BRISQUE) [49], for these metrics a smaller score indicates better perceptual quality. Table 1 presents these quality metrics for each of the data subsets. One can notice the high similarity in the quality measures between the generated visual-like images and the original visual ones (target of generation). This is an indication of the success of the generated images in imitating the quality of the targeted original images.

Table 1: Mean values (± standard deviation) of quality metrics for each subset of experimental data.

	GCF	Sharpness	NIQE	BRISQUE
oV	8.13(±0.49)	0.060(±0.007)	7.03(±1.69)	47.44(±1.49)
gV	8.19(±0.54)	0.062(±0.006)	6.35(±0.49)	48.03(±0.30)
oT	7.32(±1.43)	0.091(±0.018)	4.49 (±0.45)	33.25(±4.85)

4.1. Correlation to original images

It is very important to assessing the similarity between original and generated images. For this purpose, inter-spectral correlation coefficient (r) is used, and it can be described by:

$$r_{o:g} = \frac{\sum_{i=1}^{n}(O_i - \overline{O})(G_i - \overline{G})}{\sqrt{\sum_{i=1}^{n}(O_i - \overline{O})^2 \sum_{i=1}^{n}(G_i - \overline{G})^2}}, \quad (2)$$

where O_i denotes the i-th pixel value of the original image, and G_i corresponds to the value of the i-th pixel in the generated image. $\overline{O}$, $\overline{G}$ are mean pixel values. In addition, the correlation coefficient between original thermal images and original images taken in visible light is calculated. The determination of the correlation coefficient was preceded by the alignment of the hand position on the images using the SURF features [50] and applying geometric transformation.

Fig.2 presents boxplot of correlation coefficient between different images types. Correlation coefficient between original visible images and original thermal images indicates fair similarity, median $r_{oV:oT}$ equals 0.78. Slightly higher scores were obtained for original thermal images and generated visible-light images, however, it should be noted that gV are generated from oT. We can observe high similarity between original and generated visible images. This is another indication of the successful transformation from thermal images to generated visual images using our proposed approach.

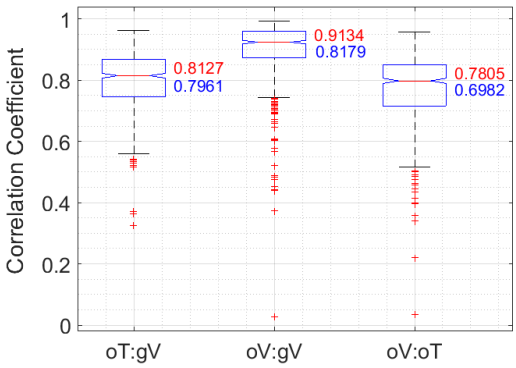


Figure 2: Boxplots of correlation coefficient between original thermal and generated visible light images (oT:gV), original and generated visible light (oV:gV), and original visible light and original thermal images (oV:oT). Median values are shown in red, whereas means are in blue.

5. Palm recognition

The evaluation of biometric recognition accuracy was based on two approaches: *feature engineering* using texture descriptor and *feature learning* employing convolutional neural networks. All processed images were resized to 224×224 pixels for both approaches, the texture-based and for the CNN-based to fit the network input size. No alignment was performed on the images.

Texture-based approach: Local Binary Patterns (LBP) [51] is one of the most popular texture descriptors used in image analysis. It represents the contrast differences in the neighbourhood of each pixel in the image. LBP descriptor values are given according to the equation:

$$LBP_{N,R}(I_C) = \sum_{n=1}^{N} s(I_n - I_C)2^{n-1}, \quad (3)$$

where N, R are the number of surrounding neighbors and the radius, respectively, I_C – central pixel, I_N – N-th pixel from neighborhood of central pixel, $s(I_n - I_C) = 1$, while $I_n - I_C > 0$ and $s(I_n - I_C) = 0$ otherwise.

The obtained values are formed as a histogram for the whole image (no-patches), as well as concatenated vector from vectors representing histograms from image patches. These patches are either 28×28 or 56×56 pixels. Feature vectors were compared using χ^2 metrics. The three versions of the LBP feature descriptors produced close performances, with the no-patch approach slightly more accurate, which can be explained by the alignment variations. Therefore, the results are only reported for the no-patch approach.

CNN-based approach: Due to their high representation learning capabilities, solutions based on convolutional neural networks are becoming increasingly popular. In this study, we employed pre-trained VGG-19 model [42] trained on the ImageNet database [43]. This network consist of 16 convolutional layers and 3 fully connected layers, constituting a classifier. This structure is modified on the classifier level to fit to our task, and fine-tuned with palm images obtained in visible light, thermal images, and generated images, respectively. Training details are presented in Tab. 2.

Table 2: VGG-19 fine-tuning details.

Options	modified VGG-19
training/validation/test	60:20:20
batch size	8
data augmentation	random rotation: $\pm 60^0$, random translation: ± 30 pixels, random reflection
optimization method	SGD (with momentum=0.9)
learning rate	0.0001
other	data shuffling before each training epoch
	L2 regularization
	validation-stop

6. Experimental results

Texture-based approach: To evaluate verification accuracy we created 10 session-disjoint splits. The averaged Receiver Operating Characteristic (ROC) curves with mean equal error rates (EERs) are plotted in Fig.3.

Comparison of LBP features of the same images type: oT, oV and gV gives EER respectively: 35.47%, 31.01% and 24.43%. It can be concluded that the proposed transformation approach has allowed to extract static features of thermal images that are more resistant to temperature changes over time and less sensitive to external and physiological factors.

Comparisons between thermal images and generated or original visible images produce EER of around 50% (random), while a relatively low error rate is found for comparisons between original and generated visual images, which is 33.87% on average. This accuracy is close to the verification accuracy of the inter-spectrum original visual images (31.01% EER) and thus proves the high success of our transformation approach for cross-spectral comparisons.

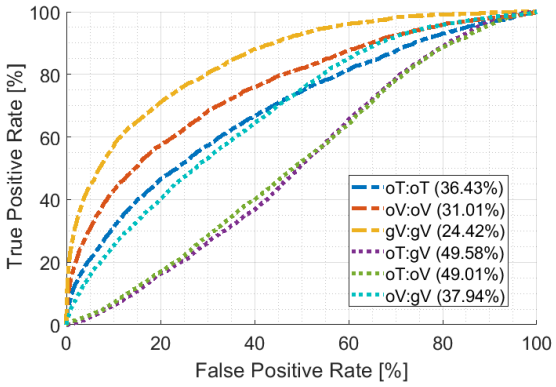


Figure 3: ROC curves obtained for method based on LBP features. Dash-dot lines: comparisons within one type of images. Dotted lines: comparison between different types of images. EER values are presented in brackets.

CNN-based approach: Similarly to the analysis of LBP features, a 10-fold division of session-disjoint data was made, however, the data was divided into three parts: training/validation/test in a ratio of 6:2:2. Averaged ROC curves with mean EERs are plotted in Fig.4.

For comparisons between images of the same type, the EER is about 1.6% for oV and gV, and for thermal images over 3%. The verification error between the thermal images was reduced from 3.04% to 1.65% after the proposed transformation into the visual space (oV:oV). This proves our goals in being more robust to thermal variations and taking advantage of visual-trained networks after our transformation. For the cross-spectral verification, our proposed transformation improved the accuracy from 37.12% EER (oT:oV) down to 16.25% (oV:gV).

7. Conclusions

The aim of this article was to compare thermal images in the visual domain, as well as the cross-spectral comparison of thermal and visual palm images. To achieve that, we successfully proposed a novel approach to transfer thermal images into the visual domain. This allows to take advantage of the high performing deep networks pre-trained on images from that domain (visual). We built a solution that targets the alignment-free and scale-invariant nature of palm images, as well as the limited size of the training data. This was achieved by utilizing the CRN trained with a loss func-

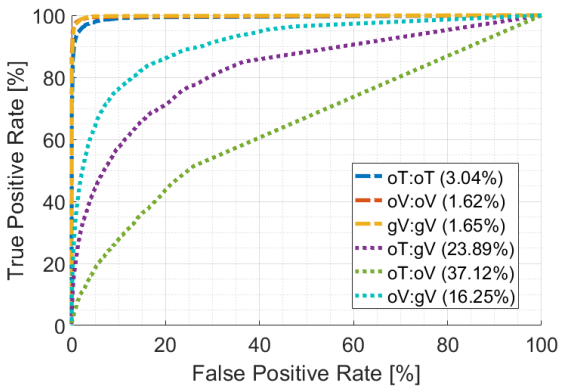


Figure 4: Similar to Fig. 3, but ROC curves obtained for CNN-based approach.

tion based on the contextual similarity. We showed that the proposed transformation produced images of similar quality as the original visual images, characterized by high correlation to them. We used these generated images to perform thermal-to-thermal and thermal-to-visual verification. This was achieved by using hand-crafted features, as well as embedding generated by pre-trained CNN. The comparison of generated images from thermal-to-thermal comparisons lead to higher verification accuracy in comparison to using the original thermal images, achieving 24.42% and 1.65% EER in comparison to 36.43% and 3.04% when comparing the thermal images directly using LBP and CNN-based features, respectively. The EER of cross-spectrum verification using CNN-extracted features was reduced from 37.12% to 16.25% after transferring the thermal image into the visual domain. It must be noted that using thermal images for verification helps avoiding presentation attacks. However, an attacker may transfer a thermal image to the visual-spectrum and attack a visual verification system.

Acknowledgement

This work was 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.

References

[1] F. Galton, *Finger Prints*. Macmillan and Company, 1892.

[2] L. Stasiak, "Support vector machine for hand geometry-based identity verification system," *Proc.SPIE, Photonics Applications in Astronomy, Communications, Industry, and High-Energy Physics Experiments*, vol. 6347, 2006.

[3] E. Yoruk, E. Konukoglu, B. Sankur, and J. Darbon, "Shape-based hand recognition," *IEEE Transactions on Image Processing*, vol. 15, 2006.

[4] D. Zhang, W. Kong, J. You, and M. Wong, "Online palmprint identification," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 25, 2003.

[5] K. Tiwari, C. J. Hwang, and P. Gupta, "A palmprint based recognition system for smartphone," *Future Technologies Conference (FTC), San Francisco, CA, USA*, 2017.

[6] L. Fei, G. Lu, W. Jia, S. Teng, and D. Zhang, "Feature Extraction Methods for Palmprint Recognition: A Survey and Evaluation," *IEEE Transactions on Systems, Man, and Cybernetics: Systems*, vol. PP, 2018.

[7] A. Kumar, "Toward More Accurate Matching of Contactless Palmprint Images Under Less Constrained Environments," *IEEE Transactions on Information Forensics and Security*, vol. 14, pp. 34–47, 2018.

[8] J. Hashimoto, "Finger Vein Authentication Technology and Its Future," *Symposium on VLSI Circuits, Digest of Technical Papers., Honolulu, USA*, 2006.

[9] Y. Zhou and A. Kumar, "Human Identification Using Palm-Vein Images," *IEEE Transactions on Information Forensics and Security*, vol. 6, 2011.

[10] H. Wan, L. Chen, H. Song, and J. Yang, "Dorsal hand vein recognition based on convolutional neural networks," *IEEE International Conference on Bioinformatics and Biomedicine (BIBM)*, 2011.

[11] N. Damer, F. Maul, and C. Busch, "Multi-biometric continuous authentication: A trust model for an asynchronous system," in *19th International Conference on Information Fusion, FUSION 2016, Heidelberg, Germany, July 5-8, 2016*. IEEE, 2016, pp. 2192–2199.

[12] E. Bartuzi, K. Roszczewska, A. Czajka, and A. Pacut, "Unconstrained Biometric Recognition Based on Thermal Hand Images," in *2018 International Workshop on Biometrics and Forensics (IWBF)*, June 2018, pp. 1–8.

[13] E. Bartuzi and M. Trokielewicz, "Thermal Features for Presentation Attack Detection in Hand Biometrics," *IEEE Int. Conference on Biometrics: Theory Applications and Systems (BTAS), Los Angeles, USA*, 2018.

[14] A. Konga, D. Zhangb, and M. Kamelc, "A survey of palmprint recognition," *Elsevier Science Inc, Pattern Recognition*, vol. 42, pp. 1408–1418, 2009.

[15] J. Patil and C. Nayak and M. Jain, "Palmprint recognition using DWT, DCT and PCA techniques ," 2015.

[16] D. Tamrakar and P. Khanna, "Palmprint verification using competitive index with PCA ," 2011.

[17] W. El-Tarhouni and L. Boubchir and N. Al-Maadeed and Mosa Elbendak and A. Bouridane, "Multispectral palmprint recognition based on local binary pattern histogram Fourier features and Gabor filter," 2016.

[18] J. Almaghtuf and F. Khelifi, "Self-geometric relationship filter for efficient SIFT key-points matching in full and partial palmprint recognition ," 2018.

[19] Q. Zhao and X. Wu and W. Bu, "Contactless palmprint verification based on SIFT and iterative RANSAC," 2013.

[20] W. Jia and R. Hu and Y. Lei and Y. Zhao and J. Gui, "Histogram of Oriented Lines for Palmprint Recognition," 2013.

[21] A. Kong and D. Zhang, "Competitive coding scheme for palmprint verification," 2004.

[22] Q. Zhao and W. Bu and X. Wu, "Sift-based image alignment for contactless palmprint verification," 2013.

[23] A. Czajka and P. Bulwan, "Biometric verification based on hand thermal images," in *2013 International Conference on Biometrics (ICB)*, June 2013, pp. 1–6.

[24] R. K. Rowe, U. Uludag, M. Demirkus, S. Parthasaradhi, and A. K. Jain, "A multispectral whole-hand biometric authentication system," *Biometrics Symp., 2007, MD, USA*, 2007.

[25] S. Samoil and S. N. Yanushkevich, "Depth assisted palm region extraction using the kinect v2 sensor," *Emerging Security Technologies (EST), 2015 6th Int. Con., Germany*, 2015.

[26] ——, "Multispectral hand recognition using the kinect v2 sensor," *Evolutionary Computation (CEC), Canada*, 2016.

[27] Y. Hao, Z. Sun, T. Tan, and C. Ren, "Multispectral palm image fusion for accurate contact-free palmprint recognition," *15th IEEE International Conference on Image Processing, San Diego, CA, USA*, 2008.

[28] M. Faundez-Zanuy, J. Mekyska, and X. Font-Aragones, "A new hand image database simultaneously acquired in visible, near-infrared and thermal spectrums," *Cognitive Computation*, vol. 6, p. 230240, 2014.

[29] ——, "Thermal hand image segmentation for biometric recognition," *IEEE Aero El Sys Mag*, vol. 28, pp. 4–14, 2013.

[30] M. Sarfraz and R. Stiefelhagen, "Deep perceptual mapping for thermal to visible face recognition," *International Journal of Computer Vision*, pp. 1–11, 2015.

[31] 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, CA, USA*. IEEE, 2008, pp. 465–468.

[32] B. S. Riggan, N. J. Short, and S. Hu, "Thermal to visible synthesis of face images using multiple regions," in *2018 IEEE Winter Conference on Applications of Computer Vision (WACV)*, March 2018, pp. 30–38.

[33] 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)*, Feb 2018, pp. 174–181.

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

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

[36] J.-Y. Zhu, T. Park, P. Isola, and A. A. Efros, "Unpaired image-to-image translation using cycle-consistent adversarial networks," *arXiv preprint*, 2017.

[37] 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*. IEEE Computer Society, 2017, pp. 5967–5976.

[38] Q. Chen and V. Koltun, "Photographic image synthesis with cascaded refinement networks," in *IEEE International Conference on Computer Vision, ICCV 2017, Venice, Italy*. IEEE Computer Society, pp. 1520–1529.

[39] 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, Quebec, Canada*, pp. 1790–1798.

[40] 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*. IEEE Computer Society, 2016, pp. 2414–2423.

[41] 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*, vol. 11218. Springer, 2018, pp. 800–815.

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

[43] 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.

[44] 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 Corua, Spain, 2019*. IEEE (under review), 2019.

[45] K. Matkovic, L. Neumann, A. Neumann, T. Psik, and W. Purgathofer, "Global contrast factor-a new approach to image contrast." *Computational Aesthetics*, pp. 159–168, 2005.

[46] X. Gao, S. Z. Li, R. Liu, and P. Zhang, "Standardization of face image sample quality," in *International Conference on Biometrics*. Springer, 2007, pp. 242–251.

[47] P. Wasnik, K. B. Raja, R. Ramachandra, and C. Busch, "Assessing face image quality for smartphone based face recognition system," in *Biometrics and Forensics (IWBF), 2017 5th International Workshop on*. IEEE, 2017, pp. 1–6.

[48] A. Mittal, R. Soundararajan, and A. C. Bovik, "Making a "completely blind" image quality analyzer," *IEEE Signal Process. Lett.*, vol. 20, no. 3, pp. 209–212, 2013.

[49] A. Mittal, A. Moorthy, and A. Bovik, "Blind/referenceless image spatial quality evaluator," in *Conference Record of the 45th Asilomar Conference on Signals, Systems and Computers, ACSCC 2011, CA, USA*. IEEE, pp. 723–727.

[50] H. Bay, A. Ess, T. Tuytelaars, and L. V. Gool., "Surf:speeded up robust features," *Computer Vision and Image Understanding (CVIU)*, vol. 110, pp. 346–359, 2008.

[51] T. Ojala, M. Pietikainen, and T. Maenpaa, "Multiresolution gray scale and rotation invariant texture classification with local binary patterns," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 24, pp. 971–987, 2002.