Thermal Features for Presentation Attack Detection in Hand Biometrics

Ewelina Bartuzi
Biometrics and Machine Intelligence Lab
Research and Academic Computer Network
Kolska 12, 01-145 Warsaw, Poland
ewelina.bartuzi@nask.pl

Mateusz Trokielewicz
Institute of Control and Computation Engineering
Warsaw University of Technology
Nowowiejska 15/19, 00-665 Warsaw, Poland
m.trokielewicz@elka.pw.edu.pl

Abstract

This paper proposes a method for utilizing thermal features of the hand for the purpose of presentation attack detection (PAD) that can be employed in a hand biometrics system's pipeline. By envisaging two different operational modes of our system, and by employing a DCNN-based classifiers fine-tuned with a dataset of real and fake hand representations captured in both visible and thermal spectrum, we were able to bring two important deliverables. First, a PAD method operating in an open-set mode, capable of correctly discerning 100% of fake thermal samples, achieving Attack Presentation Classification Error Rate (APCER) and Bona-Fide Presentation Classification Error Rate (BPCER) equal to 0%, which can be easily implemented into any existing system as a separate component. Second, a hand biometrics system operating in a closed-set mode, that has PAD built right into the recognition pipeline, and operating simultaneously with the user-wise classification, achieving rank-1 recognition accuracy of up to 99.75%. We also show that thermal images of the human hand, in addition to liveness features they carry, can also improve classification accuracy of a biometric system, when coupled with visible light images. To follow the reproducibility guidelines and to stimulate further research in this area, we share the trained model weights, source codes, and a newly created dataset of fake hand representations with interested researchers.

1. Introduction

Personal features of the hand have been employed for the authentication of humans since the early days of modern biometrics, in the form of fingerprint minutiae introduced as early as 1892 by Galton [1], geometric features [2, 3], palmprints [4, 5, 6, 7], and finger and hand vein patterns [8, 9, 10]. Recently, thermal features of the hand have gained some attention in the biometric community with the work of Bartuzi *et al.* [11], showing that heat distributions carry discriminatory information and allow to build a biometric method based solely on the thermal hand representations.

Imaging of the human hand for biometric purposes is rather straightforward, and in cases when only texture features, such as principal lines or minutiae of the palmprint are used, does not require a specialized equipment. However, what makes it easy to collect a biometric sample, usually also reveals a presentation attack (PA) vulnerability, which may involve presenting the system with a fake representation of the hand, such as a paper printout, prosthetics, displays, or using a genuine hand in a non-conformant scenario (*e.g.*, use under coercion, or presentation improper enough to compromise the system).

Thus, an important piece in every well-designed biometric system's pipeline is a way to mitigate such attempts, *i.e.*, presentation attack detection (PAD). Although thermal features of the hand have been used as cues for determining a person's identity, we are not aware of any PAD method that would employ such traits for counteracting fake representation attacks. This paper thus offers the following **contributions to the state-of-the-art:**

- a presentation attack detection method employing thermal features, using a static image of the hand, based on a deep convolutional neural network (DCNN) model trained in both identity-driven (closed set) and authenticity-driven (open-set) approaches,
- a dataset of fake hand representations collected in the visible and thermal infrared wavelength rages, complementary to the existing dataset of genuine hand samples from the *MobiBits* database,
- trained model weights and source codes.

Model weights, source codes and a complementary dataset of fake visible light and thermal infrared hand representations can be obtained at: `http://zbum.ia.pw.edu.pl/EN/node/46`.

978-1-5386-7180-1/18/$31.00

This article is laid out as follows. In Sec. 2 we review the existing PAD methods for alleviating spoof attacks in palmprint and dorsal hand vein biometrics. Sec. 3 describes a subset of the *MobiBits* database used for the purpose of this study and details the process of creating a complementary set of fake hand representations for training and evaluating our PAD method, which is introduced in Sec. 4. Finally, experimental results are reported and discussed in Sec. 5, whereas Sec. 6 provides relevant conclusions.

2. Related work

Kanhangad and Kumar explored local binary pattern features for discerning genuine palmprint samples from paper printouts, reaching over 97% accuracy [12]. Kanhangad *et al.* pointed to a high, almost 80% risk of accepting fake palmprint samples as genuine ones using one of the existing academic methods, and proposed an anti-spoofing mechanism that analyzes the reflectance of the palmprint samples, and is said to be able to correctly classify 99% of the genuine samples, paper printouts, and computer display photographs into either *real* or *fake* subsets [13]. A method for detecting presentation attacks in dorsal hand-vein biometrics is introduced by Bhilare *et al.* in [14], employing a histogram of oriented gradients performed on LoG-filtered images and an SVM with majority voting for image classification, reaching EER from 0.16% to 0.8%. A fusion of texture-based approaches and image quality assessment for face and palmprint PAD is introduced by Farmanbar and Toygar in [15], and tested on several publicly available datasets of face and palmprint samples. Bhilare *et al.* followed up on their earlier work in [16], introduing a spoof sample database – 'PALMspoof', and a PAD method that is said to outperform their earlier LBP-based approach by 12.73 percentage points in classification error rate.

All of the methods reviewed above employ hand-crafted texture features or image statistics of visible light or near infrared images for determining PAD cues. However, to our knowledge, there are no prior papers or published research that would employ thermal imaging for the purpose of presentation attack detection of hand samples, despite this being perhaps the most natural choice for building a PAD method, as replicating the exact heat distribution of the human hand seems a dubious and difficult task for an attacker to perform.

3. Experimental data

3.1. Dataset of visible light and thermal hand images

For the purpose of this work we use a subset of the *MobiBits*, a multimodal mobile biometric database including images of palm side of the hand [17]. The data were collected from 53 volunteers (20 female and 33 male, both hands, bringing a total of 106 classes). The age of subjects ranged from 14 to 71 years. The data were collected in three different acquisition sessions organized during three following months, twice in each session, approximately 15 minutes apart. All samples were acquired in typical office conditions with air conditioning on and ambient temperature set to 24^oC.

The discussed subset, containing **hand images** can be categorized into two sample types:

1. **RGB** – images taken with a rear camera of the CAT s60 mobile phone (480×640 pixels) in three sessions: with no temperature influence, after warming with a hot pillow, and after cooling with a cold compress. Additionally, measurement included unconstrained acquisition (raised hand) and acquisition stabilized by a glass stand.

2. **TH** – thermal images, taken simultaneously with the images in visible light, constituting the fourth layer of the image: RGB + TH, but with lower resolution of 240×320 pixels).

3.2. Creating fake hand representations

As a part of the experimentation performed in this work, we have extended the *MobiBits–Hand* subset with fake samples (two images per class). Hand images were printed and then photographed in a similar way as real hands were when creating the database, and with the use of a same CAT s60 device. Heat distribution of hand was imitated by the hand of a living human placed under the printout during data collection to make the presentation attack more plausible. Example hand images from the MobiBits-Hand subset together with their respective fake representations are shown in Fig. 1.

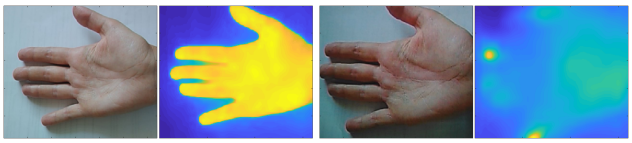


Figure 1. **From left to right:** an example real hand image in visible light, under thermal imaging, and respective fake representations.

4. Proposed methodology

4.1. Model architecture

Two popular DCNN architectures were used in our experiments, namely the VGG-19 model introduced by Simonyan [18], and a shallower AlexNet model proposed by Krizhevsky [19]. VGG-19, called 'very deep' at the time of publication, consists of 16 convolutional layers – which serve as feature extractor, followed by three fully connected (FC) layers, constituting a classifier. AlexNet, on the other

Table 1. Summary of the experimental protocol in each scenario (SPE - seconds per epoch).

Mode / DCNN model	**AlexNet**	**VGG-19**
Authenticity-driven	**binary classification** **open-set** (subject-disjoint) **2 images per subject** 200 real and 200 fake images	
	RGB: 12 epochs TH: 11 epochs	RGB: 10 epochs TH: 7 epochs
Identity-driven	**class-wise prediction** 106 identity classes + 1 class of fake representations **closed-set** (sample-disjoint) **45 images per class**	
	RGB: 45 epochs TH: 75 epochs	RGB: 25 epochs TH: 36 epochs

hand, comprises only 5 convolutional layers, followed by a similar, three-layer fully-connected classifier. Both networks participated in the ILSVRC competition [20] and are pre-trained on the ImageNet database [21] for the task of classifying natural images.

These pre-trained models were then altered by modifying the bottleneck layers of the classifier stage to fit the task investigated in this paper, and fine-tuned with a database of genuine and fake, visible light and thermal hand images. These experiments were performed twofold, namely in an **authenticity-driven open-set** scenario, and **identity-driven closed-set** scenario. These are described in detail in the following sections.

4.2. Authenticity-driven mode

First, the DCNNs were used as binary classifiers, yielding a decision related to the authenticity of the sample being processed by the network, regardless of its claimed identity. The result here is assigning the sample either a *real* or *fake* label, together with a probability score obtained from one of the two output softmax neurons of the model.

The DCNN operating in this mode can serve as an additional security element in any biometric system's pipeline, without introducing modifications to its architecture.

4.3. Identity-driven mode

In the second part of the experiments reported in this paper, the DCNNs were fitted to operate in a closed-set scenario, where detection of fake hand representations is carried out simultaneously with providing a class-wise prediction of the probe sample. The last FC layers of each model were modified so that the number of output softmax neurons is $N + 1$, where N is the number of classes, and the additional neuron represents the *fake* class. If a sample is classified as genuine, a typical prediction of the target class is given, together with a probability score obtained from the softmax layer. If, however, a fake is detected, then the sample is assigned a *fake* label with a probability score.

For a biometric system expected to operate in a closed-set mode only, this simplifies the integration of the PAD component into the recognition pipeline with little additional cost.

4.4. Training and evaluation

For the network training and testing procedure, 10 subject-disjoint train/validation/test data splits were created. They were made with replacement, making them statistically independent and allowing us to assess the variance of the estimated error rates. The network was then trained with each train subset independently for each split, with the training being automatically stopped after achieving a non-increasing accuracy on the validation subset with patience of 10, and evaluated on the corresponding test subset. The training was performed with stochastic gradient descent as the minimization method with momentum $m = 0.9$ and learning rate of 0.0001, with the data being passed through the network in mini batches of 16 images. Additionally, the training data were shuffled before each training epoch.

For the authenticity-driven, open-set mode, the data were divided in a subject-disjoint manner, so that classes chosen randomly for each of the train/validation/test splits do not overlap subject-wise. For the identity driven closed-set mode, subject-disjoint division is not possible, therefore for the second experiment the samples for each of the train/validation/test subsets were chosen randomly from each class to approximately meet the 60:20:20 criterion. Table 1 summarizes details of the experiments performed for each network and each scenario.

In each mode, the networks were trained separately with RGB (visible light) and TH (thermal) images. Then, a score fusion at the softmax level was performed, so that in each mode three score distributions for each DCNN model were available:

- scores obtained from RGB images only,
- scores obtained from TH images only,
- scores obtained by averaging the above scores.

5. Results and discussion

5.1. Authenticity-driven mode

Figure 2 presents the accuracy of classifying hand representations into either the *real* or the *fake* class, for both the AlexNet and VGG-19 models. Both classifiers achieve perfect accuracy when confronted with thermal samples, but even with only the visible light samples, we can still expect a very high, more than 98% classification accuracy provided by the larger VGG-19, and slightly less than 96% obtained from the much shallower AlexNet model. Since employing thermal features yield a perfect accuracy alone, we do not report on the score fusion results here.

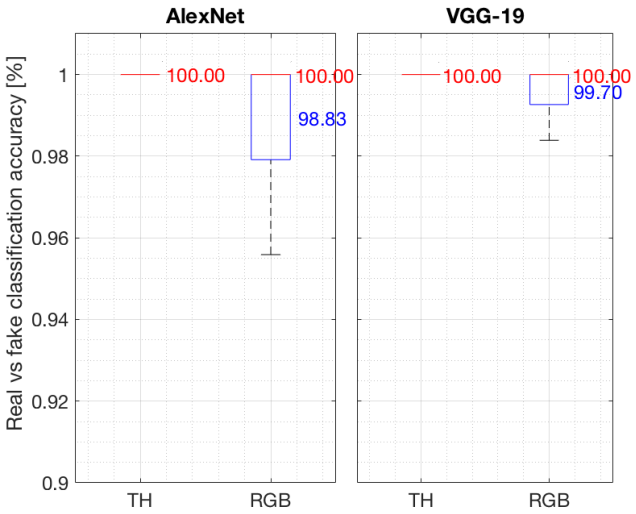


Figure 2. Boxplots representing differences in accuracy of classification into real and fake classes for thermal (TH) and visible light (RGB) hand representations for two DCNN models. Median values are shown in red, whereas mean values are shown in blue. Height of each box denotes an inter-quartile range (IQR), spanning from the first (Q1) to the third (Q3) quartile, whereas whiskers span from $Q1 - 1.5 \cdot IQR$ to $Q3 + 1.5 \cdot IQR$.

5.2. Identity-driven mode

Classification accuracies obtained from the experiments performed using the identity-driven operational mode, in which the networks were given the task of not only detecting *fake* representations, but also classifying the hand sample into one of the *real* classes in a closed-set scenario, are shown in Fig. 3.

In addition to classification accuracies, we also report the distributions of the scores used to generate boxplots shown in Fig. 3. These are shown in Figures 4 and 5, for scores obtained using the AlexNet and VGG-19 models, respectively. Notably, whereas some overlapping of the bins denoting scores corresponding to *genuine* and *fake* comparisons when RGB and TH images are used separately, this is almost completely rectified when a score fusion by arithmetical averaging is performed.

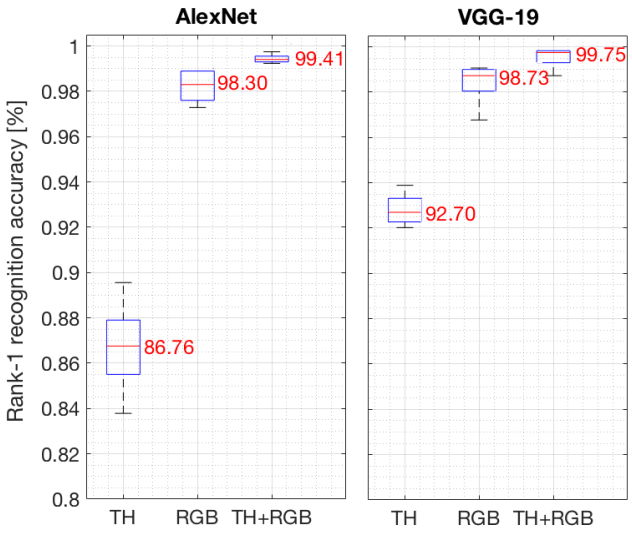
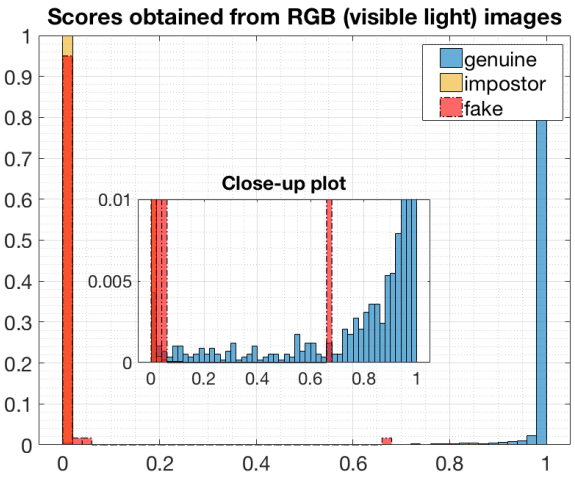


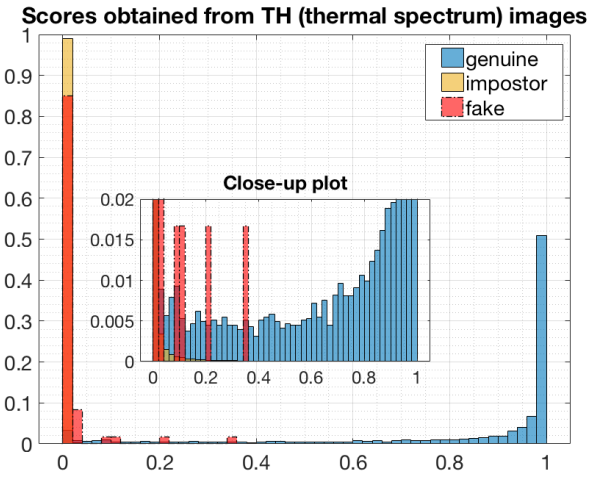
Figure 3. Same as in Fig. 2, but showing accuracies for the identity-driven operational mode on a closed set of *real* identities complemented by a *fake* class. Scores obtained from thermal (TH) and visible light (RGB) hand representations, and scores obtained by averaging the TH and RGB scores (TH+RGB).

6. Conclusions

This paper is the first known to us work that explores thermal features of the hand for the purpose of assessing sample liveness for presentation attack detection, and employs deep convolutional neural networks adapted to two different operational scenarios and fine-tuned with a dataset of real and fake representations of hands imaged in both visible light and thermal spectrum.

For the authenticity-driven mode, in which the algorithm's task is to discern *fake* representations from *real* ones, we were able to achieve perfect, 100% accuracy averaged over 10 subject-disjoint, statistically independent train/validation/test data splits when thermal images are used with both the AlexNet and VGG-19 models. This translates to $APCER = 0\%$ and $BPCER = 0\%$. Also, surprisingly good performance can be expected for visible light images as well, with mean *real* vs *fake* classification accuracies equaling 98.83% and 99.70% for the AlexNet and VGG-19, respectively, averaged over the same training and testing procedure. These in turn allow to obtain low $APCER = 0.87\%$ and $BPCER = 0.55\%$ for the AlexNet, and $APCER = 0.29\%$ and $BPCER = 0.97\%$ for VGG-19. Since the much shallower AlexNet model reaches the same perfect performance with thermal images, we may hazard a guess that further optimization of the network architecture can bring down the computational cost, while maintaining the same 100% accuracy.





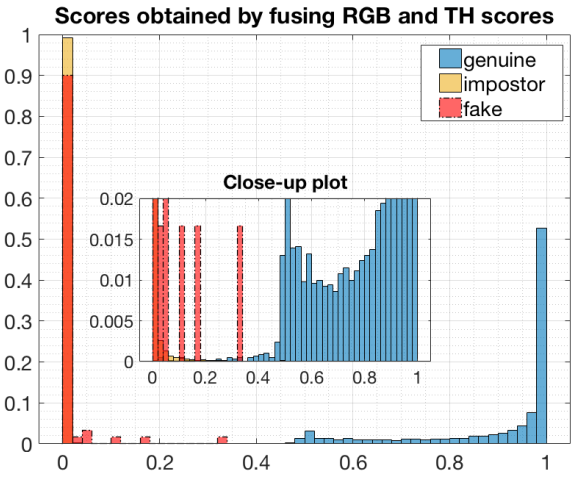
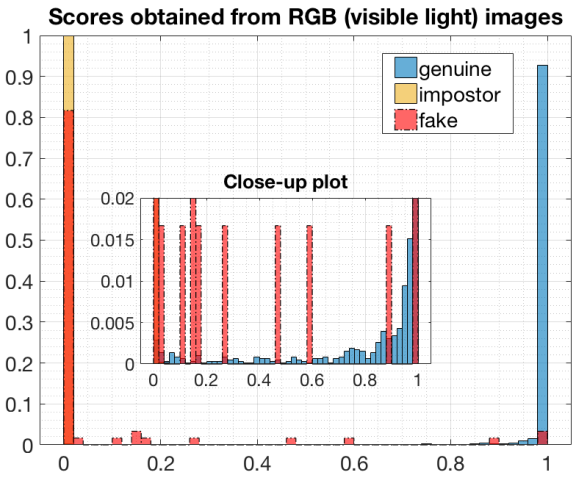
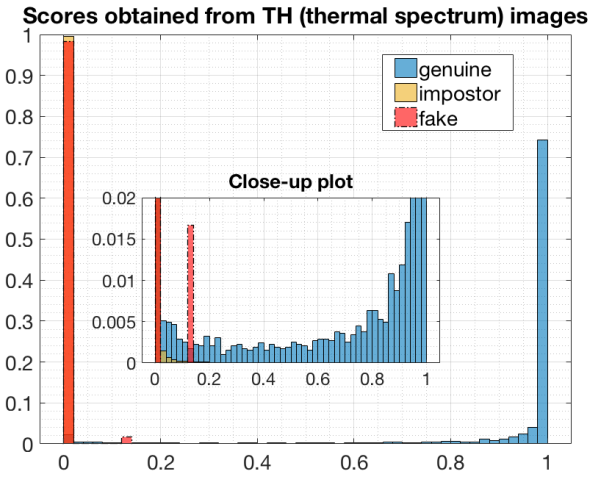


Figure 4. Score distributions generated with the AlexNet model for visible light images, thermal images, and a fusion of both. Smaller plots in the middle of each larger plot is a close-up of the lower registers of the ordinate axis.





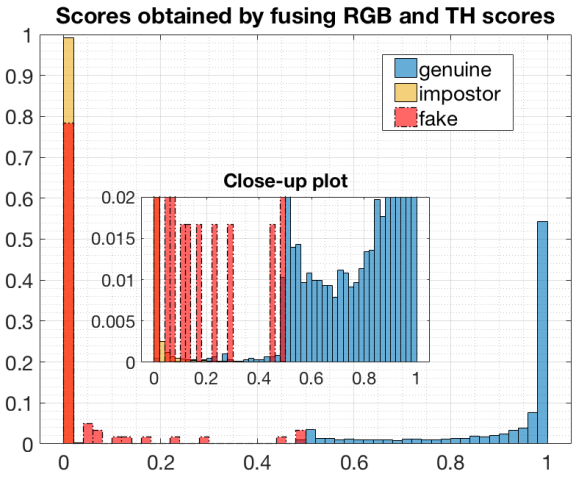


Figure 5. Same as in Fig. 4, but for the VGG-19 model.

In the second operational scenario, namely the identity-driven closed-set mode, the solutions presented in this paper were able to offer 99.41% and 99.75% rank-1 recognition accuracy on the set of identities joined by a class of fake

representations, for AlexNet and VGG-19, respectively. Although the VGG-19 model allows for a slightly better accuracy than AlexNet, it is also much more computationally complex. Therefore, here as well we may argue that with further architecture optimizations, most of the recognition performance accuracy can be retained, while significantly reducing the cost of the solution.

A second interesting observation that this experiment delivers, is that thermal features alone carry enough personal information to allow for over 92% rank-1 recognition accuracy, and, when coupled with visible light features, are able to raise the overall performance of the system to an almost ideal accuracy of 99.75%. Thus, **utilizing thermal features of the human hand appears not only to be a perfect method for a robust presentation attack detection method, but also a way to improve the overall performance of the biometrics system**, provided that both types of data are collected at both the enrollment, and the verification stages.

This work follows the IEEE guidelines for research reproducibility by offering the following contributions together with the paper: a) trained DCNN model weights and example source codes, and b) a dataset of fake hand representations that is complementary to the corresponding subset of the *MobiBits* database [17].

References

[1] F. Galton, *Finger Prints*. Macmillan and Company, 1892. [Online]. Available: https://books.google.pl/books?id=MgJ0tQEACAAJ

[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] 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.

[12] V. Kanhangad and A. Kumar, "Securing palmprint authentication systems using spoof detection approach," *Proc.SPIE*, vol. 9067, pp. 9067 – 9067 – 5, 2013. [Online]. Available: https://doi.org/10.1117/12.2051724

[13] V. Kanhangad, S. Bhilare, P. Garg, P. Singh, and N. Chaudhari, "Anti-spoofing for display and print attacks on palmprint verification systems," 2015. [Online]. Available: https://doi.org/10.1117/12.2180333

[14] S. Bhilare, V. Kanhangad, and N. Chaudhari, "Histogram of oriented gradients based presentation attack detection in dorsal hand-vein biometric system," in *2017 Fifteenth IAPR International Conference on Machine Vision Applications (MVA)*, May 2017, pp. 39–42.

[15] M. Farmanbar and Ö. Toygar, "Spoof detection on face and palmprint biometrics," *Signal, Image and Video Processing*, vol. 11, no. 7, pp. 1253–1260, Oct 2017. [Online]. Available: https://doi.org/10.1007/s11760-017-1082-y

[16] S. Bhilare, V. Kanhangad, and N. Chaudhari, "A study on vulnerability and presentation attack detection in palmprint verification system," *Pattern Analysis and Applications*, vol. 21, pp. 769–782, 2018.

[17] E. Bartuzi, K. Roszczewska, M. Trokielewicz, and R. Białobrzeski, "Mobibits: Multimodal Mobile Biometric Database," *17th International Conference of the Biometrics Special Interest Group (BIOSIG2018)*, 2018.

[18] K. Simonyan and A. Zisserman, "Very Deep Convolutional Networks for Large-Scale Image Recognition," 2014. [Online]. Available: https://arxiv.org/abs/1409.1556

[19] A. Krizhevsky, I. Sutskever, and G. E. Hinton, "ImageNet Classification with Deep Convolutional Neural Networks," pp. 1097–1105, http://papers.nips.cc/paper/4824-imagenet-classification-with-deep-convolutional-neural-networks.pdf, 2012.

[20] O. Russakovsky, J. Deng, H. Su, J. Krause, S. Satheesh, S. Ma, Z. Huang, A. Karpathy, A. Khosla, M. Bernstein, A. C. Berg, and L. Fei-Fei, "ImageNet Large Scale Visual Recognition Challenge," *International Journal of Computer Vision (IJCV)*, vol. 115, no. 3, pp. 211–252, 2015.

[21] 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.