On the Impact of Different Fabrication Materials on Fingerprint Presentation Attack Detection

Lázaro J. González-Soler[1], Marta Gomez-Barrero[2], Leonardo Chang[3], Airel Pérez Suárez[1], Christoph Busch[2]
[1]Advanced Technologies Application Center, La Habana, Cuba
ljsoler24@gmail.com, asuarez@cenatav.co.cu
[2] da/sec - Biometrics and Internet Security Research Group, Hochschule Darmstadt, Germany
{marta.gomez-barrero,christoph.busch}@h-da.de
[3]Tecnologico de Monterrey, School of Engineering and Science, Mexico
lchang@tec.mx

Abstract

Presentation Attack Detection (PAD) is the task of determining whether a sample stems from a live subject (bona fide presentation) or from an artificial replica (Presentation Attack Instrument, PAI). Several PAD approaches have shown high effectiveness to successfully identify PAIs when the materials used for the fabrication of these PAIs are known a priori. However, most of these PAD methods do not take into account the characteristics of PAIs' species in order to generalise to new, realistic and more challenging scenarios, where materials might be unknown. Based on that fact, in this work, we explore the impact of different PAI species, fabricated with different materials, on several local-based descriptors combined with the Fisher Vector feature encoding, in order to increase the robustness to unknown attacks. The experimental results over the well-established benchmarks of the LivDet 2011, LivDet 2013 and LivDet 2015 competitions reported error rates outperforming the top state-of-the-art in the presence of unknown attacks. Moreover, the evaluation revealed the differences in the detection performance due to the variability between the PAI species.

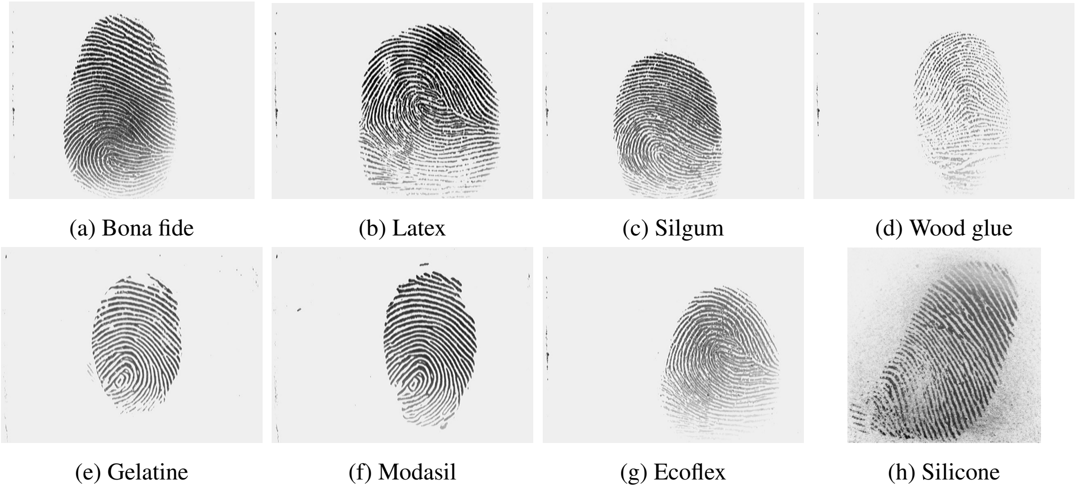
1. Introduction

Fingerprint-based biometric systems have experienced a huge development in the last decades. This is due to their user convenience and their ability to provide a higher security level than traditional password or token based authentication systems. However, fingerprint recognition schemes are still vulnerable to presentation attacks, where an attacker tampers with the sensor with an artificial replica of a fingerprint (i.e., a Presentation Attack Instrument, PAI). Such PAIs can be fabricated with a wide range of materials, including, but not limited to, gelatine, silicone, silgum, wood glue, or latex. Some samples are shown in Fig. 1.

To tackle the security concerns posed by the aforementioned attacks, several presentation attack detection (PAD) methods have been proposed over the last years in order to determine whether a sample stems from a live subject (i.e. it is a bona fide presentation (BF)) or from an artificial replica (i.e., it is a presentation attack (PA)). However, even if considerable efforts have been directed to the creation of new fingerprint PAI species and their automatic detection [7, 11], it is likely that in a nearby future new materials, unknown to current PAD approaches, can be used for the fabrication of more challenging PAIs. As a consequence, the development of new PAD methods, robust to unknown PAI species, is mandatory.

In this context, Rattani *et al.* [18] analysed the characteristics of several spoof materials taken from the LivDet 2011 Biometrika dataset, in order to improve the detection of PAIs built with unknown materials (i.e., enhance the detection capabilities for unknown attacks). According to the authors, PAIs from Biometrika 2011 dataset, which are built using different spoof materials, exhibit differences in the surface coarseness and in the contrast between ridges and valleys. Their evaluation over the LivDet 2011 database showed that the use of different fabrication materials leads to different degrees of generalisation ability for the PAD method. In addition, the use of adaptive Weibull-calibrated support vector machines (W-SVMs) for classification yielded an improvement of up to 44% in terms of performance generalisation.

More recently, Gonzalez-Soler *et al.* proposed in [8] a new method based on the Bag of Words (BoW) encoding of local keypoint based descriptors, in order to enhance the generalisation capabilities of the local features to unknown attacks. On their evaluation over the LivDet 2011 database,



(a) Bona fide (b) Latex (c) Silgum (d) Wood glue

(e) Gelatine (f) Modasil (g) Ecoflex (h) Silicone

Figure 1: Samples from a bona fide and several PAIs fabricated with different materials.

they reported an average classification error rate under 5% (i.e., a relative improvement of 14% with respect to the state-of-the-art). However, no analysis was carried out on the impact of different PAI species on the detection capabilities of the proposed methodology.

In spite of those and other valuable works, one main question remains unanswered: what is the impact of the different PAI species of the PAD performance? In other words, to which extent are some PAI species harder to detect? And does this difficulty vary for different feature extractors and for different sensors? On the other hand, is there a relationship between the average detection performance of a particular PAD approach and its robustness to different PAI species?

To the best of our knowledge, very few works have addressed these issues. Only at the end of 2018 Chugh and Jain analysed 12 different PAI species over a databased acquired with a single CrossMatch sensor in [4]. In particular, they grouped the features extracted by the Fingerprint Spoof Buster PAD method [3] in order to derive a training set, comprising only six PAI species, which achieved a similar detection performance to the classifier trained with all PAI species.

Therefore, to answer the aforementioned questions over a publicly available dataset, we study in this article both continuous and binary descriptors in combination with the Fisher Fector (FV) feature encoding for fingerprint PAD purposes. This approach builds upon the encoding scheme presented in [8], increasing the detection performance with the use of FV encoding instead of BoW. In more details, we explore the impact of several materials commonly used for the fabrication of the PAIs on the detection performance of several descriptors in combination with FV: a dense version of Scale Invariant Feature Transform, denoted dense-SIFT (DSIFT) [13], Binarized Statistic Invariant Features (BSIF) [12], Local Binary Pattern (LBP) [16], Histogram of Oriented Gradients (HOG) [5], Speed-Up Robust Features (SURF) [1], Binary Robust Independent Elementary Features (BRIEF) [2], and Oriented FAST and Rotated BRIEF (ORB) [19]. All these descriptors have reported remarkable results in several computer vision tasks.

The detection performance of the proposed PAD approach is evaluated over the LivDet 2011 [23], LivDet 2013 [6] and LivDet 2015 [14] databases, in order to allow the reproducibility of the results and a fair benchmark with the available literature. A detailed analysis of the performance for each local descriptor and PAI species is included in order to answer the questions posed above. Furthermore, we compare the performance achieved with the state-of-the-art results reported over the LivDet databases in [3], following the standard evaluation protocol described in [15].

The remainder of this paper is organised as follows: in Sect. 2, we describe the proposed PAD method. The experimental evaluation and the results are discussed in Sect. 3. Finally, conclusions and future work directions are presented in Sect. 4.

2. Presentation Attack Detection Method

As mentioned in Sect. 1, the proposed PAD method relies on local descriptors. These descriptors are subsequently encoded following the Fisher Vector (FV) approach in order to determine the best discriminating common feature space, and thereby allowing a better generalisation to unknown scenarios. Fig. 2 shows an overview of the proposal, which consists on the following three main steps: *i)* local descriptors extraction, both binary and continuous; *ii)* decorrelated local feature encoding via FV; and *iii)* classification into bona fide or attack presentation using a linear SVM.

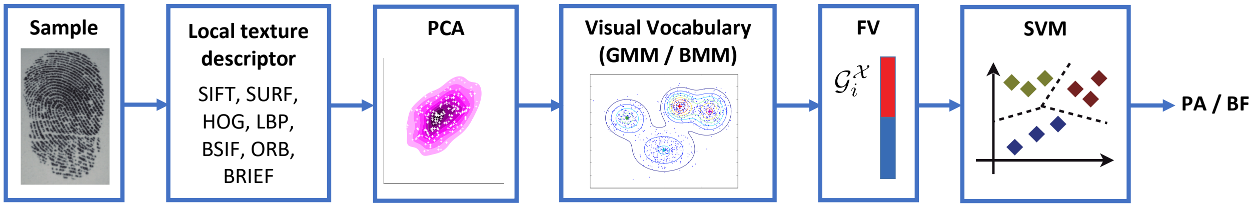

Sample
Local texture descriptor
SIFT, SURF, HOG, LBP, BSIF, ORB, BRIEF
PCA
Visual Vocabulary (GMM / BMM)
FV
$\mathcal{G}_i^{\mathcal{X}}$
SVM
PA / BF

Figure 2: PAD approach overview. First, local features are computed at different scales. Then this features are encoded using a previously learned visual vocabulary. Finally, the fingerprint descriptor is classified using a linear SVM.

In the recent past, the FV representation has received a considerable attention due to its effectiveness for image classification and retrieval [20, 22]. This approach derives a kernel from a generative model of the data. Then, it characterises a particular local features patch by its deviation from the generative model computed in the first step (e.g., Gaussian Mixture Models (GMMs) [20] for continuous data, or Bernoulli Mixture Models (BMMs) [22] for binary data). Both binary and continuous features are studied in this work. For the corresponding detailed descriptions, the reader is referred to [13, 12, 16, 5, 1, 2, 19].

For the continuous descriptors (i.e., DSIFT, SURF, PHOG, LBP, and BSIF), we follow the pipeline proposed in [17], training a GMM model with diagonal covariances from the decorrelated local features extracted in the previous step. In general, a GMM, which models the generation process of continuous local descriptors in any image, is known as a probabilistic visual vocabulary [17], whose K-components are represented by the mixture weights (w_k), Gaussian means (μ_k) and covariance matrix (σ_k), with $k = 1, \dots, K$. This leads to an image representation, which captures the average first-order statistics and second-order differences between the local features and each of the GMM centres [21]:

$$\phi_k^1 = \frac{1}{N\sqrt{w_G^k}} \sum_{i=1}^{N} \alpha_i(k) \left(\frac{x_i - \mu_k}{\sigma_k} \right), \tag{1}$$

$$\phi_k^2 = \frac{1}{N\sqrt{2w_G^k}} \sum_{i=1}^{N} \alpha_i(k) \left(\frac{(x_i - \mu_G^k)^2}{\sigma_k^2} - 1 \right), \tag{2}$$

where $\alpha_i(k)$ is the soft assignment weights of the i-th feature x_i to the k-th Gaussian. It is important to highlight that w_G^k, μ_G^k and σ_k are computed during the training stage. Finally, the FV representation that defines a fingerprint image is obtained by stacking those differences: $\phi_G = \left[\phi_1^{(1)}, \phi_1^{(2)}, \cdots, \phi_K^{(1)}, \phi_K^{(2)}\right]$.

With the aim of clustering the extracted local features with GMM diagonal covariances, those features are decorrelated using Principal Component Analysis (PCA) [10]. In our approach, the local descriptor dimension was reduced to d = 64 components, hence resulting the FV representation in a $2Kd$ size vector, where K is the number of Gaussian components in the GMM, and d is the dimension of local descriptors. It should be noted that local descriptors such as BSIF, PHOG and LBP lead to 256-dimensional feature vectors, which were reduced to 64 components.

On the other hand, for encoding binary features (i.e., ORB and BRIEF), we follow a pipeline in [22] and we train a Binary Mixture Model (BMM), whose K-components are represented by the correspoding weights (w_B^k) and means (μ_B^k), with $k = 1, \dots, K$. Thus, a closed-form approximation of FV representation is given by [22]:

$$\phi_{\mu_{kd}} = \left(\frac{1}{T} \sum_{t=1}^{T} \gamma_k(x_t) \frac{(-1)^{1-x_t^d}}{\mu_{kd}^{x_t^d}(1-\mu_{kd})^{1-x_t^d}} \right) F_{kd}^{-\frac{1}{2}}, \tag{3}$$

where

$$\gamma_k(x_t) = \frac{w_k p_k(x_t|\theta|)}{\sum_{k=1}^{K} w_k p_k(x_t|\theta|)} \tag{4}$$

$$F_{kd} = T w_k \left(\frac{\sum_{k=1}^{K} w_k \mu_{kd}}{\mu_{kd}^2} + \frac{\sum_{k=1}^{K} w_k (1-\mu_{kd})}{(1-\mu_{kd})^2} \right) \tag{5}$$

It is important to highlight that the FV representation based on BMM approach only takes into account the gradients with respect to μ_{kd}, and that it does not require data decorrelation (i.e., PCA is not used).Therefore, the KD-dimensional FV representation is defined as: $\phi_B = [\phi_{\mu_{kd}}]$, with $k = 1, \dots, K$ and $d = 1, \dots, D$. This way, the FV representation based on BMM approach defines a compact vector, whose size is half of the GMM-based FV representation.

3. Experimental Evaluation

3.1. Experimental Protocol

The experiments were conducted on the well-established and freely available benchmarks of the LivDet 2011 [23], LivDet 2013 [6] and LivDet 2015 [14] competitions (see Table 1 for a description). It should be noted that we only used three out of four datasets from the LivDet 2013 benchmark: Biometrika, Italdata and Swipe. In particular, the Crossmatch dataset was not used for PAD evaluation because, according to the LivDet competition organisers [7],

Table 1: PAI species in the LivDet databases.

DB	Dataset	PAI species
LivDet 2011	Biometrika	EcoFlex, Silgum,
	Digital P.	Gelatine, Latex,
	Italdata	WoodGlue, PlayDoh,
	Sagem	Silicone
LivDet 2013	Biometrika	EcoFlex, Gelatine,
	Italdata	Latex, Modasil, BodyDouble
	Swipe	WoodGlue, PlayDoh
LivDet 2015	GreenBit	BodyDouble, EcoFlex
	Digital P.	PlayDoh, Gelatine,
	Biometrika	Latex,
	Crossmatch	WoodGlue

there are several anomalies in the fingerprint data from that dataset, which was accordingly deprecated.

With the main aim of evaluating the impact of different PAI species built using different spoof materials on the PAD performance, the experimental evaluation has a twofold objective: *i)* analyse the impact of different PAI species on the proposed method, following the standard LivDet evaluation protocol (see Sect. 3.2.1); and *ii)* compare the performance and robustness to unknown attacks of the proposed approach to the top state-of-the-art methods, following the experimental protocol proposed in [15] (see Sect. 3.2.2).

In order to stablish a fair benchmark with the current literature, we first measure the detection accuracy in terms of the Average Classification Error Rate (ACER), which is defined as:

$$\text{ACER}(\delta) = \frac{\text{APCER}(\delta) + \text{BPCER}(\delta)}{2} \tag{6}$$

where Attack Presentation Classification Error Rate (APCER) is the percentage of misclassified presentation attacks for a fixed threshold, and Bona fide Presentation Classification Error Rate (BPCER) is the percentage of misclassified bona fide presentations for a fixed threshold. These last metrics are defined within the recent ISO/IEC 30107-3 standard for the evaluation of PAD apporaches [9], and will also be reported in terms of the corresponding Detection Error Trade-Off (DET) curves and the Detection Equal Error Rate (D-EER) (i.e., operating point where APCER = BPCER). This allows a rigorous analysis for different operation points or values of δ. Finally, the BPCER for a fixed APCER of 10% (BPCER10), 5% (BPCER20), and 1% (BPCER100) is reported.

3.2. Experimental Results

3.2.1 Robustness to Different PAI Species

The first set of experiments evaluates the impact of different spoof materials included in the LivDet databases on the PAD method described in Sect. 2. To that end, all testing and training images are acquired using the same sensor. Moreover, the same PAI species are used for training and testing in order not to bias the results. The detection performance is analysed in terms of the D-EER, which is computed individually for each PAI species and descriptor.

The mean and standard deviation (std) of the D-EERs for each descriptor is shown with a boxplot in Fig. 3a. As it can be observed, among all local descriptors, DSIFT yields the lowest error rates across different materials (i.e., a mean D-EER of 1.88%). In addition, the standard deviation (i.e., size of the corresponding box) is the lowest (0.96%), thereby indicating its higher robustness to different PAI species with respect to the other descriptors considered.

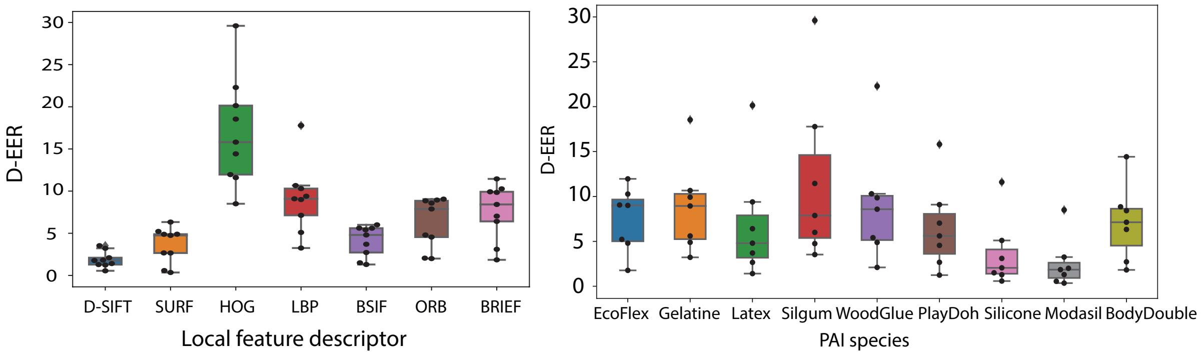
Following DSIFT, the BSIF-based encoding achieves a D-EER of 4.07%, which is slightly higher than the one achieved by the SURF-based encoding (D-EER = 3.59%). However, the SURF's std (2.13%) is higher than the one yielded by BSIF (std = 1.84). Therefore, we may conclude that BSIF is more robust to material variability than SURF. In addition, we can answer one of the questions posed in Sect. 1: there is no direct relationship between the overall performance of a particular PAD method and its robustness to different PAI species, even if these are known during training.

Now, Fig. 3b shows the D-EER distributions per PAI species or fabrication material. As it can be observed, the highest variability is yielded by the PAIs made with Silgium (mean = 11.57% and std = 9.30%), thereby reflecting its high resemblance with the bona fide samples (see Figs. 1a and 1c). On the other hand, the modasil PAIs (Fig. 1f) showed a very distinct appearance, with no noise in the ridges as in the bona fide samples, and are consequently easier to detect by all descriptors. As a consequence, modasil yields the lowest mean D-EER and std (mean = 2.54% and std = 2.80%). To sum up, we can conclude that, as it could be expected, some PAI species (e.g., silgium) are harder to detect than others (e.g., modasil).

3.2.2 Robustness to Unknown PAI Species

For the second set of experiments, we have selected the DSIFT-based encoding, which achieved the best error rates on the previous analysis. In order to evaluate its robustness to unkown attacks, we follow the experimental protocol described in [15]. Table 2 reports a benchmark in terms of ACER against the top state-of-the-art method presented in [3].

As it can be observed, the DSIFT-based encoding yields an average ACER of 3.03%, which is slightly worse than the one reported by [3] (ACER = 2.93%). However, our proposal shows a better detection performance for three out

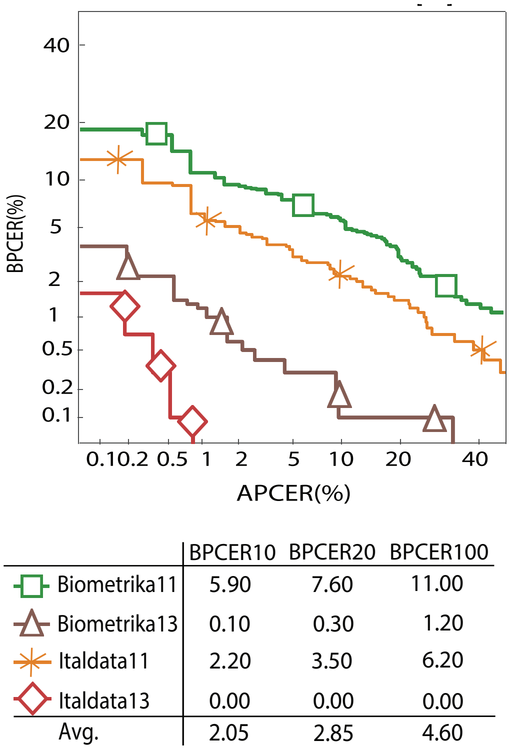


(a) (b)

Figure 3: **Evaluation of different PAI species**. D-EER values per (a) local feature descriptor, and (b) PAI material.

Table 2: **Comparison with the state-of-the-art**. ACER of the proposed approach, compared to the results reported in [3], following the experiemental protocol described in [15].

Dataset	**Training set**	**Test set**	**DSIFT-based encoding**	**[3]**
Biometrika 2011	EcoFlex, Gelatine, Latex	Silgum, Woodglue	7.05	**4.60**
Biometrika 2013	Modasil, Woodglue	EcoFlex, Gelatine, Latex	**1.00**	1.30
Italdata 2011	EcoFlex, Gelatine, Latex	Silgum, Woodglue, Other	**3.78**	5.20
Italdata 2013	Modasil, Woodglue	EcoFlex, Gelatine, Latex	**0.30**	0.60
		Avg.	3.03	**2.93**



	BPCER10	BPCER20	BPCER100
□ Biometrika11	5.90	7.60	11.00
△ Biometrika13	0.10	0.30	1.20
✳ Italdata11	2.20	3.50	6.20
◇ Italdata13	0.00	0.00	0.00
Avg.	2.05	2.85	4.60

Figure 4: **ISO-compilant performance evaluation** of the D-SIFT + FV encoding approach following the experimental protocol proposed by [15].

of the four datasets. More specifically, the DSIFT-based encoding reports on these three dataset an average ACER of 1.69%, which outperforms the top state-of-the-art [3] (ACER = 2.37%) by a relative 28.69%. It should be noted that our selected proposal yields high error rates in those test datasets including PAIs built with Silgum (see Fig. 3b). Despite that, the proposed method is able to successfully detect new PAI species, which were fabricated with unknown materials, thereby indicating its high capacity to generalise to new, realistic and more challenging scenarios.

Finally, we evaluate the performance of our proposal in compliance with the ISO/IEC 30107-3 [9], thereby allowing a deeper performance analysis at different operating points. The results are depicted in Fig. 4. As it can be observed, the low error rates show the main strength of FV encoding: regardless of the dataset and even in the presence of unknown PAI species, a low BPCER can be achieved for different APCERs. More specifically, for Italdata13, a BPCER of 0% can be achieved for any APCER $\geq$ 0.01%, and for Biometrika13, a low BPCER100 of 1.20% is achieved for an APCER of 1%. In particular, for an APCER of 1%, both Italdata11 (BPCER100 = 6.20%) and Italdata13 (BPCER = 0.00%) outperform the results yielded by the top state-of-the-art [3] (BPCER100 = 7.89% and BPCER100 = 0.68%, respectively). These results suggest that the DSIFT-based proposal is able to be used in real applications demanding high security levels.

4. Conclusions

In this work, a new PAD method based on FV encoding and several continuous and binary local descriptors was proposed. Experimental results conducted over spoof materials taken from different LivDet competitions showed that the DSIFT-based encoding is more robust to PAIs fabricated using different spoof materials than other descriptors. More specifically, this approach yielded an average D-EER of 1.88% and a standard deviation of 0.96%, which indicate its high capacity to correctly identify PAIs built with different materials.

Furthermore, a benchmark with the top state-of-the-art [3] shows its high capacity to successfully detect presentation attacks fabricated with unknown materials. Even for high security scenarios (i.e., APCER = 1%), an average BPCER of 4.60% can be achieved, thereby allowing for high user convenience at the same time.

Building upon the knowledge gained in this study, we will focus on the development of fingerprint PAD methods which can better generalise to unknown materials, evaluating different feature encoding approaches, and also considering deep learning based approaches.

Acknowledgments

This work was partially 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 the Center for Research in Security and Privacy (CRISP). This work was carried out during an internship of L. J. González-Soler at da/sec.

References

[1] H. Bay, T. Tuytelaars, and L. Van Gool. SURF: Speeded up robust features. In *Proc. ECCV*, pages 404–417, 2006.

[2] M. Calonder, V. Lepetit, C. Strecha, and P. Fua. BRIEF: Binary robust independent elementary features. In *Proc. ECCV*, pages 778–792, 2010.

[3] T. Chugh, K. Cao, and A. K. Jain. Fingerprint spoof buster: Use of minutiae-centered patches. *IEEE Trans. on Information Forensics and Security*, 2018.

[4] T. Chugh and A. K. Jain. Fingerprint presentation attack detection: Generalization and efficiency. *arXiv preprint arXiv:1812.11574*, 2018.

[5] N. Dalal and B. Triggs. Histograms of oriented gradients for human detection. In *Proc. CVPR*, pages 886–893, 2005.

[6] L. Ghiani, D. Yambay, V. Mura, et al. LivDet 2013 fingerprint liveness detection competition 2013. In *Proc. ICB*, 2013.

[7] L. Ghiani, D. A. Yambay, V. Mura, et al. Review of the fingerprint liveness detection (LivDet) competition series: 2009 to 2015. *Image and Vision Computing*, 58:110–128, 2017.

[8] L. J. González-Soler, L. Chang, J. Hernández-Palancar, et al. Fingerprint presentation attack detection method based on a bag-of-words approach. In *Proc. CIARP*, pages 263–271, 2017.

[9] ISO/IEC JTC1 SC37 Biometrics. *ISO/IEC FDIS 30107-3. Information Technology - Biometric presentation attack detection - Part 3: Testing and Reporting*, 2017.

[10] H. Jegou, F. Perronnin, M. Douze, et al. Aggregating local image descriptors into compact codes. *IEEE Trans. on Pattern Analysis and Machine Intelligence*, 34(9):1704–1716, 2012.

[11] O. Kanich, M. Drahanskỳ, and M. Mézl. Use of creative materials for fingerprint spoofs. In *Proc. IWBF*, 2018.

[12] J. Kannala and E. Rahtu. BSIF: Binarized statistical image features. In *Proc. ICPR*, pages 1363–1366, 2012.

[13] D. G. Lowe. Distinctive image features from scale-invariant keypoints. *Int. Journal of Computer Vision*, 60(2):91–110, 2004.

[14] V. Mura, L. Ghiani, G. L. Marcialis, et al. LivDet 2015 fingerprint liveness detection competition 2015. In *Proc. BTAS*, 2015.

[15] R. F. Nogueira, R. de Alencar Lotufo, and R. C. Machado. Fingerprint liveness detection using convolutional neural networks. *IEEE Trans. on Information Forensics and Security*, 11(6):1206–1213, 2016.

[16] T. Ojala, M. Pietikäinen, and T. Mäenpää. Multiresolution gray-scale and rotation invariant texture classification with local binary patterns. *IEEE Trans. on Pattern Analysis and Machine Intelligence*, 24(7):971–987, 2002.

[17] F. Perronnin, J. Sánchez, and T. Mensink. Improving the fisher kernel for large-scale image classification. In *Proc. ECCV*, pages 143–156, 2010.

[18] A. Rattani, W. J. Scheirer, and A. Ross. Open set fingerprint spoof detection across novel fabrication materials. *IEEE Transactions on Information Forensics and Security*, 10(11):2447–2460, 2015.

[19] E. Rublee, V. Rabaud, K. Konolige, and G. Bradski. ORB: An efficient alternative to SIFT or SURF. In *Proc. ICCV*, pages 2564–2571, 2011.

[20] J. Sánchez, F. Perronnin, T. Mensink, and J. Verbeek. Image classification with the fisher vector: Theory and practice. *Int. Journal on Computer Vision*, 105(3):222–245, 2013.

[21] K. Simonyan, O. M. Parkhi, A. Vedaldi, and A. Zisserman. Fisher vector faces in the wild. In *BMVC*, volume 2, page 4, 2013.

[22] Y. Uchida, S. Sakazawa, and S. Satoh. Image retrieval with fisher vectors of binary features. *ITE Trans. on Media Technology and Applications*, 4(4):326–336, 2016.

[23] D. Yambay, L. Ghiani, P. Denti, et al. LivDet 2011-fingerprint liveness detection competition 2011. In *Proc. ICB*, pages 208–215, 2012.