IEEE Access
Multidisciplinary | Rapid Review | Open Access Journal

Received December 4, 2020, accepted December 28, 2020, date of publication January 1, 2021, date of current version January 12, 2021.

Digital Object Identifier 10.1109/ACCESS.2020.3048756

Fingerprint Presentation Attack Detection Based on Local Features Encoding for Unknown Attacks

LÁZARO JANIER GONZÁLEZ-SOLER[1], MARTA GOMEZ-BARRERO[2], (Member, IEEE), LEONARDO CHANG[3], AIREL PÉREZ-SUÁREZ[4], AND CHRISTOPH BUSCH[1], (Senior Member, IEEE)

[1]da/sec—Biometrics and Internet Security Research Group, Hochschule Darmstadt, 64295 Darmstadt, Germany
[2]Department of Computer Science, Hochschule Ansbach, 91522 Ansbach, Germany
[3]School of Engineering and Science, Tecnologico de Monterrey, Monterrey 64849, Mexico
[4]Advanced Technologies Application Center (CENATAV), Havana 12200, Cuba

Corresponding author: Lázaro Janier González-Soler (lazaro-janier.gonzalez-soler@h-da.de)

This work was supported in part by the DFG-ANR RESPECT Project under Grant 406880674, in part by the German Federal Ministry of Education and Research, and in part by the Hessian Ministry of Higher Education, Research, Science and the Arts within their joint support of the National Research Center for Applied Cybersecurity ATHENE.

ABSTRACT Fingerprint-based biometric systems have experienced a large development in the past. In spite of many advantages, they are still vulnerable to attack presentations (APs). Therefore, the task of determining whether a sample stems from a live subject (i.e., bona fide) or from an artificial replica is a mandatory requirement which has recently received a considerable attention. Nowadays, when the materials for the fabrication of the Presentation Attack Instruments (PAIs) have been used to train the Presentation Attack Detection (PAD) methods, the PAIs can be successfully identified in most cases. However, current PAD methods still face difficulties detecting PAIs built from unknown materials and/or unknown recepies, or acquired using different capture devices. To tackle this issue, we propose a new PAD technique based on three image representation approaches combining local and global information of the fingerprint. By transforming these representations into a common feature space, we can correctly discriminate bona fide from attack presentations in the aforementioned scenarios. The experimental evaluation of our proposal over the LivDet 2011 to 2019 databases, yielded error rates outperforming the top state-of-the-art results by up to 72% in the most challenging scenarios. In addition, the best representation achieved the best results in the LivDet 2019 competition (overall accuracy of 96.17%).

INDEX TERMS Local feature encoding, presentation attack detection, fingerprint, probabilistic visual vocabulary, visual vocabulary.

I. INTRODUCTION

Biometric recognition is based on the observation of distinctive anatomical and behavioural characteristics to automatically recognise a subject [1]. Among other biometric characteristics, fingerprints offer a high recognition accuracy and at the same time enjoy a high popular acceptance. Despite of these advantages, fingerprint-based recognition systems can be circumvented by launching Attack Presentations (APs), in which an artificial fingerprint, denoted as Presentation Attack Instrument (PAI), is presented to a capture device [2]–[5].

The associate editor coordinating the review of this manuscript and approving it for publication was Pengcheng Liu.

We would like to highlight that the threat posed by PAIs is not reduced to an academic issue. In 2000 Zwiesele *et al.* [6] reported that they fooled three commercial fingerprint capture devices with PAIs made of india rubber. Then, two years later Matsumoto *et al.* [4] analysed the vulnerabilities of eleven commercial fingerprint-based biometric systems to gummy fingerprints. The experimental evaluation showed that 68% to 100% of the PAIs built with cooperative methods were accepted as bona fide presentations (i.e., genuine or live fingers). In 2009, Japan reported the detection of presentation attacks in one of its airports, and in 2013, a Brazilian doctor used artificial silicone fingerprints to tamper a biometric attendance system at the Sao Paulo hospital [7].

In order to tackle those severe security threats, the development of Presentation Attack Detection (PAD) techniques,

TABLE 1. Summary of the studies focused on fingerprint PAD generalisation.

Study	Approach	Known-env	Unknown PAIs	Cross-sensor	Cross-DB
Rattani *et al.* [8]	Weibull-calibrated SVM	-	D-EER = 19.70%	-	-
Ding & Ross [9]	Ensemble of several one-class SVMs	-	D-EER = 17.60%	-	-
Nogueira *et al.*(VGG) [10]	Feasibility of three CNNs for fingerprint PAD generalisation	ACER = 3.87%	ACER = 6.30%	ACER = 19.80%	ACER = 30.70%
Pala & Bhanu (TripleNet) [11]	Triple CNNs over random extracted patches	ACER = 2.41%	ACER = 5.86%	ACER = 25.25%	ACER = 15.20%
Chugh & Jain (FSB-v1) [12]	Inception-v2 trained for classifying minutiae-centred local patches	ACER = 1.70%	ACER = 3.50%	ACER = 16.60%	ACER = 18.90%
Chugh & Jain (FSB-v2) [13]	MobileNet trained for classifying minutiae-centred local patches	ACER = 1.11%	ACER = 2.93%	ACER = 14.59%	ACER = 17.91%
Chugh & Jain [14]	Selection of the most feasible PAI species to cover the deep feature space	-	BPCER500 = 24.76%	-	-
Engelsma & Jain [15]	Ensemble of several GANs	-	BPCER500 = 50.20%	-	-
Gajawada *et al.* [16]	Universal Material Translator (UMT) to generate synthetic PAIs, thereby improving PAD generalisation	-	BPCER1000 = 21.96%	-	-
Chugh & Jain (FSG) [17]	Universal Material Generator (UMG) to style transfer between known PAI species to improve its generalisation capability	-	BPCER1000 = 8.22% Avg. Acc. = 95.88	BPCER1000 = 56.77% Avg. Acc. = 80.63%	-
Park *et al.* (TinyFCN) [18]	Lightweight CNN on the fire module of the SqueezeNet	ACER = 1.43%	ACER = 1.90%	-	-
Proposed method	Ensemble of several encodings to define a common feature space from known PAIs	**ACER = 1.74%**	**ACER = 4.32%**	**ACER = 4.08%**	**ACER = 9.15%**

ACER: Average Classification Error Rate; D-EER: Detection Equal Error Rate; BPCER: Bona fide Presentation Classification Error Rate. These metrics are defined in Sect. IV

which automatically detect PAIs presented to the biometric capture device, is required. This area of research has attracted a lot of attention within the biometric research community not only for fingerprint systems [19], [20], but also for other biometric characteristics such as face [21] or iris [22]. These PAD methods can be widely classified as hardware- or software-based approaches. Whereas the former requires dedicated, and mostly expensive, specific hardware, software-based approaches focus on dynamic or static properties and features extracted from the same biometric samples used for recognition purposes. Therefore, software-based methods are less expensive, and will be the focus of this work.

The newest fingerprint PAD techniques based on deep learning and textural features have shown to be a powerful tool to detect most PAIs [10]–[13], [18], [23]. However, they share a common limitation: they depend both on *i*) the material used for fabricating the PAIs, and *ii*) the capture device used for acquiring the fingerprint samples. More specifically, their error rates are increased by a factor of five to 18 times when either PAIs' materials or capture devices utilised are not known a priori (see Tab. 1).

In order to address in this work the issue of generalisation to unknown factors, we analyse the combination of local features (i.e., Scale-Invariant Feature Transform, SIFT [24]) with three different general purpose feature encoding approaches, which have shown remarkable results in object classification tasks [25]–[27]: *i*) Bag of Words (BoW), *ii*) Vector of Locally Aggregated Descriptors (VLAD), and *iii*) Fisher Vector (FV). The local descriptors, computed over the image gradient, allow capturing different geometrical artefacts produced by materials used for generating the PAIs. By assuming that unknown attacks share homogeneous features such as texture, shape, and appearance with known PAI species, the aforementioned encoding approaches assign each local descriptor (i.e., SIFT) to the closest entry in a *visual vocabulary* [28]. This visual vocabulary defines a common feature space, thereby allowing a better generalisation to unknown attacks or capture devices.

In order to evaluate the performance of the proposed methods and to allow the reproducibility of the results, we conduct a thorough experimental evaluation on the well-known LivDet 2011, LivDet 2013, LivDet 2015, and LivDet 2019 databases. The performance is reported in compliance with the ISO/IEC 30107-3 international standard on PAD evaluation [5], thereby allowing a rigorous analysis of the results. The experimental evaluation shows the capacity of the new method to be used in high security applications. In addition, we would like to highlight that the proposed method in the Fingerprint Liveness Detection Competition 2019 achieved the best detection performance with an average accuracy of 96.17% [29]. This database was designed to evaluate the PAD generalisation performance over unknown PAIs.

The remainder of this paper is organized as follows: related works are summarised in Sect. II. In Sect. III, we describe the proposed PAD methods. The experimental evaluation is presented in Sect. IV. Finally, conclusions and future work directions are presented in Sect. V.

II. RELATED WORK

The task of determining whether a sample stems from a live subject (i.e., it is a bona fide presentation - BP) or from an artificial replica (i.e., it is an attack presentation - AP) is still an open problem, which has received a considerable amount of attention in the recent past [19], [20], [30]. As it was mentioned in Sect. I, we focus on static software-based fingerprint PAD methods, since they are the most cost efficient. Those techniques can be broadly categorised as perspiration-, pore-, image quality-, and texture-based approaches. In our work, we review those texture methods built upon deep learning and addressing scenarios with unknown factors. A summary of fingerprint PAD algorithms considered in this work is reported in Tab. 1. For further details on other methods, the reader is referred to [19], [20], [30].

A. TRADITIONAL CNN-BASED TECHNIQUES

Recently, the broad advances experienced by deep learning approaches and their success in several computer vision tasks have led to the development of powerful architectures for fingerprint PAD. Those schemes have, in turn, significantly outperformed any earlier PAD techniques.

Nogueira *et al.* [10] benchmarked three classic Convolutional Neural Networks (CNNs). One of their proposals achieved the best results in the LivDet 2015 competition, with an overall accuracy of 95.5%. In spite of those promising results, the main limitation of these methods is that they learn features from a whole image with a fixed size. In many cases, within the LivDet databases, the Region of Interest (ROI) covers only a small area of the entire image (e.g., 19% for some subsets of LivDet 2011), thus not being large enough to allow an efficient PA detection. This is highlighted by the results achieved on the LivDet 2011 - Italdata dataset, where the Average Classification Error Rate (ACER) increased up to 9.2%.

In order to address the small ROI issue, Pala and Bhanu [11] proposed training a triple CNN, which is fed with a randomly extracted patch of a fixed size per image. Despite the improvement with respect to the previous holistic-image-based approach [10], its error rates showed a poor detection performance for Italdata 2011 (i.e., ACER of 5.10%). Based on the fact that PAIs produce spurious minutiae on a fingerprint image, Chugh *et al.* [12], [13] proposed a deep learning framework for independently classifying local patches around the extracted minutiae. The final BP vs. AP decision was defined as the average between PAD scores of the local patches. This approach additionally allows finding AP regions inside a sample, even if the PAI only covers part of the underlying fingerprint. The method, named Fingerprint Spoof Buster (FSB-v2) [13], achieved the lowest ACER values reported so far over the LivDet databases.

Finally, Park *et al.* proposed in [18] an efficient CNN based on the fire module of the SqueezeNet to optimise the hardware and time requirements. The experimental evaluation over the LivDet 2011 to 2015 showed that the proposed CNN outperformed, for some datasets, the FSB-v1 [12], at the same time reducing over 6 times the execution time. However, a benchmark of this PAD method against FSB-v2 [13] under more challenging unknown attack scenarios or capture devices has not been carried out yet.

To sum up, the main drawback of the aforementioned methods is their high dependency both on the PAI fabrication materials and the capture device, thereby resulting in a high accuracy decrease for challenging scenarios, as depicted Tab. 1. It should be observed that the detection performance of traditional CNN-based methods is substantially decreased by a factor of two to 10 times when PAIs at hand are: *i*) unknown in training, *ii*) acquired with an unknown capture device, or *iii*) obtained with the same capture device, yet at different acquisition conditions and years.

B. ANOMALY DETECTION-BASED TECHNIQUES

In order to tackle previous generalisation issues, several anomaly detection-based approaches built upon handcrafted features have been followed. Given that the detection of unknown PAIs can be seen like an open set recognition problem,[1] Rattani *et al.* [8] proposed an automatic adaptation of Weibull-calibrated support vector machines (SVMs) which is relatively robust for open set recognition. The experimental results for the LivDet 2011 database showed that detection equal error rates (D-EERs) oscillated between 20 and 30% in the presence of unknown PAI species. Over the LivDet 2011 dataset, Ding and Ross analysed an ensemble of one-class SVMs trained only on BP samples in [9], which lowered the error rates to 10-22% over the same dataset.

More recently, in an extension of FSB [13], Chugh and Jain [14] identified a subset of six out of 12 PAI species to cover the entire PAI deep feature space, hence yielding a detection performance similar to known attacks scenarios. That is, training the FSB with only those six PAI species and testing on all 12 species results in a BPCER = 10.24% at APCER = 0.2%, very close to the BPCER = 9.03% when eleven PAI species are used for training. In spite of these impressive results, it should be noted that the selection of the training PAI plays a crucial role in this study.

This dependency is highlighted again by Engelsma and Jain in [15], where multiple generative adversarial networks (GANs) are trained on bona fide images acquired with the RaspiReader sensor. From the same 12 PAI species, six are used for training and six for testing. In a benchmark with the method proposed in [9], the GANs outperform the SVMs. However, the average BPCERs achieved for an APCER = 0.2% vary from 31.02% to 68.58%, depending on the training set used. This shows again a high sensitivity to different training datasets. In addition, this approach is not directly comparable to those based on conventional (e.g., Crossmatch or Greenbit) capture devices, since a specific hardware, namely the RaspiReader, was used to acquire the samples.

[1] Open-set problems address the possibility of new classes during testing, that were not seen during training.

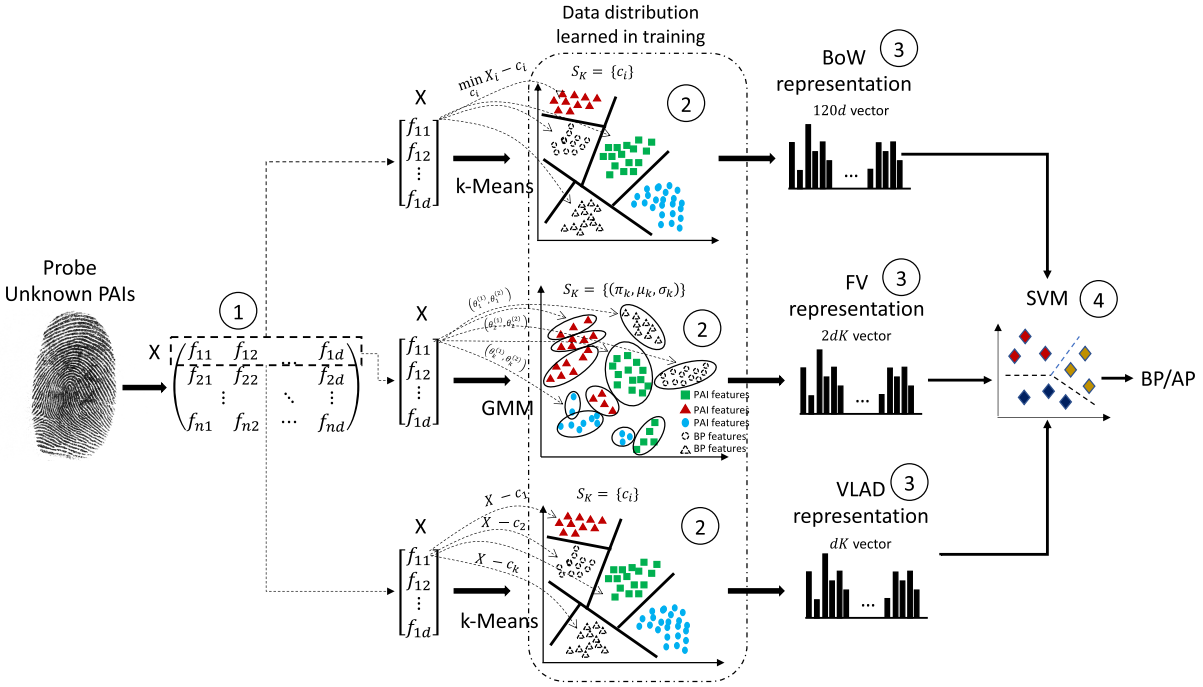


FIGURE 1. **Proposed PAD approach overview. First, dense-SIFT descriptors are computed at different scales over the whole input image. These features are subsequently encoded using a previously learned visual vocabulary by means of three different approaches: a) BoW, b) FV, and c) VLAD. The fingerprint descriptor per encoding is separately classified using a linear SVM.**

Gajawada *et al.* tried to tackle this dependency on the PAI species contained in the training set from a different perspective in [16]. They propose a so-called deep learning based "Universal Material Translator" (UMT). Given a reduced number (e.g., five) of samples from a new PAI species, the UMT extracts their main appearance features to embed them into a database of bona fide samples, in order to generate synthetic samples of the new PAI species. Those synthetic samples can then be utilised to train any CNN. Over the LivDet 2015 database, the authors showed how the proposed approach can improve up to 17% the detection rates, achieving a remarkable BPCER of 21.96% for an APCER = 0.1%. However, it should be noted that this method does require some samples (i.e., five) of the analysed unknown PAI species.

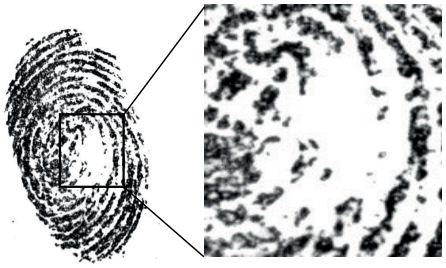
Finally, by assuming that unknown PAIs species share texture (style) information with known PAIs, Chugh and Jain [17] extended the work in [16] by combining texture styles of pre-defined PAI species to generate new synthetic unknown PAIs. Those synthetic data could, in turn, be employed as training to enhance the generalisation capability of any end-to-end PAD approach. To that end, the authors proposed an "Universal Material Generator" (UMG), which, unlike [16], required no unknown PAI species to train. The experimental protocol reported a slight detection performance improvement against the proposed baseline and LivDet 2017 winner [31] (i.e., an average accuracy of 95.88% vs. 95.44% for the baseline and 94.01% for the LivDet winner). However, the cross-sensor evaluation showed a poor detection performance (i.e., an average accuracy of 80.63%).

In this context, our method tackles the issue of detection performance degradation in absence of knowledge on relevant factors (i.e., attack species, capture devices, or databases) by transforming the local descriptors extracted from the fingerprint samples into a common feature space. Building upon that PAIs share homogeneous texture information with known PAIs and heterogeneous with those bona fide presentations, this representation allows a definition of semantic sub-groups from known samples to improve the generalisation capabilities to more challenging scenarios, not needing any samples of the unknown attacks for training.

III. PROPOSED METHOD

Fig. 1 shows an overview of the proposed PAD approach which is based on the aforementioned three different feature encoding approaches. In the first common processing step, the Pyramid Histogram of Visual Words (PHOW) [32] algorithm is used to extract local features from the whole input image: the so-called dense Scale-Invariant Feature Transform (dense-SIFT) descriptors (Sect. III-A). Subsequently, three encoding methods are applied to transform the local descriptors into a common feature space: *i*) Bag-of-Words (BoW) [25] (Sect. III-B1), *ii)* Fisher Vector (FV) [28] (Sect. III-B2), and *iii)* Vector Locally Aggregated Descriptors (VLAD) [33] (Sect. III-B3). Finally, the bona fide (BP) vs attack presentation (AP) decision for a sample at hand is taken by a linear Support Vector Machine (SVM) (Sect. III-C).

a) Bona fide ridge pattern



b) Attack presentation ridge pattern

FIGURE 2. BP vs. AP ridge patterns. a) good quality BP ridge pattern; b) AP whose ridge pattern was affected by the coarseness of the artificial material employed on its fabrication.

A. LOCAL FEATURES EXTRACTION: DENSE-SIFT DESCRIPTORS

As local feature descriptors we have chosen the dense-SIFT approach, computed over the image gradient, since they can capture lower coherence areas introduced by the coarseness of different PAI fabrication materials, as depicted in Fig. 2. In particular, the Pyramid Histogram Of visual Words (PHOW) approach proposed by [32] computes the SIFT descriptors densely at fixed points on a regular grid with uniform spacing S (e.g., 5 pixels), as illustrated in Fig. 3 (left). For each point in the grid, the dense-SIFT descriptor computes the gradient vector for each pixel in the feature point's neighbourhood (Fig. 3, top right), taking into account 8 different directions. Subsequently, a normalized 8-bin histogram of gradient directions (Fig. 3, bottom right) is built over 4×4 sample regions. In addition, in order to account for the scale variation between fingerprints, these dense-SIFT descriptors are computed over four circular patches or windows with different scales $\sigma = \{5, 7, 10, 12\}$. Therefore, each point in the grid is represented by four SIFT descriptors (i.e., one per σ) comprising a total number of 128 features (i.e., 4×4 8-bin histograms).

It should be noted that windows with different scales allow extracting local information of fingerprints at different resolution levels, thereby detecting variable-size artefacts produced in the fabrication of PAIs. In addition, near-uniform local patches do not yield stable keypoints or descriptors. Therefore, we have used a fixed threshold τ on the average norm of the local gradient in order to remove local descriptors from regions with an average norm value close to zero (i.e., low contrast regions).

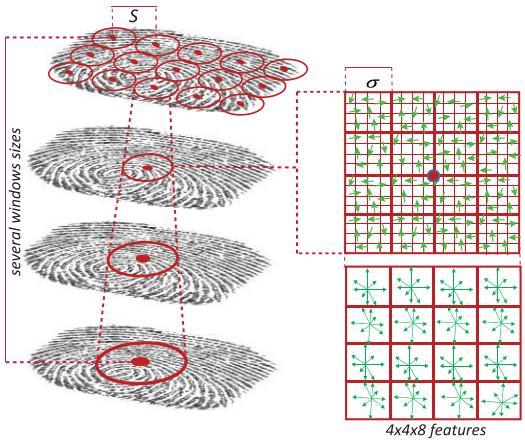


FIGURE 3. dense-SIFT descriptors computed at fixed points on a regular grid, striding with an uniform spacing S and using several scales σ.

B. LOCAL FEATURE ENCODING

In the second stage of the PAD algorithm, three different feature encoding approaches for the dense-SIFT descriptors are analysed.

1) BAG OF WORDS (BoW)

This technique was first developed for text categorization tasks, in which a text document is assigned to one or more categories based on its content [34]. To that end, BoW represents a text document by a sparse histogram of word occurrence based on a visual vocabulary. Following this idea, Csurka *et al.* [25] adopted and applied this method to represent local features from an image in terms of the so-called *visual words*. Our method builds upon this approach.

As proposed in [35], the BoW representation first computes the visual vocabulary as a codebook with K different centroids or visual words (see Fig. 1, top) with k-means clustering. Then, the BoW representation is defined as the histogram of the number of image local descriptors assigned to each visual word. Its computation is summarised in Fig. 4. An m-level pyramid of spatial histograms is used in order to incorporate spatial relationships between patches. For that purpose, the fingerprint image is partitioned into increasingly fine sub-regions, and the dense-SIFT descriptors inside each sub-region are assigned to the closest centroid among the K visual words, using a fast version of k-means clustering [36]. Subsequently, the histograms inside each sub-region are computed and transformed into a single and final feature vector by a homogeneous kernel map [37].

2) FISHER VECTOR (FV)

BoW approaches encode local features using a *hard assignment*, in which a local descriptor is only assigned to one visual word based on a similarity function. In contrast, the Fisher Vector (FV) method derives a kernel from a generative model of the data (e.g., Gaussian Mixture Model, GMM), and describes how the distribution of a set of local descriptors, extracted from unknown PAIs, differs from the known

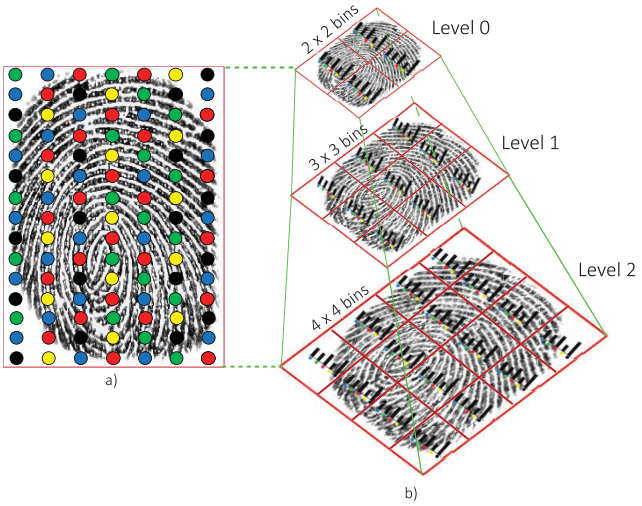


FIGURE 4. Example of pyramid of spatial histograms. a) Quantized features using *k*-means. b) 3-level pyramid of spatial histograms built from quantized features.

PAI distribution previously learned by the adopted generative model [28]. The aforementioned generative model can be understood as a *probabilistic* visual vocabulary, thereby allowing a *soft assignment*. Thus, the FV paradigm encodes not only the number of descriptors assigned to each region, but also their position in terms of their deviation with respect to the pre-defined model.

As proposed in [38], we train a GMM model with diagonal covariances from the dense-SIFT descriptors extracted on the previous step. In particular, a GMM on K-components, which is represented by their mixture weights (π_k), means (μ_k), and covariance matrices (σ_k), with $k = 1, \ldots, K$, allows discovering semantic sub-groups from known PAIs and BP samples, which could successfully enhance the detection of unknown attacks. In order to build those semantic groups, the dense-SIFT descriptors are firstly decorrelated using Principal Component Analysis (PCA) [33], hence reducing their size to $d = 64$ components while retaining 95% of the variance. Then, the FV representation which captures the average statistics first-order and second-order differences between the local features and each semantic sub-groups previously learnt by the GMM is computed [39].

Let $\mathbf{X}$ be a local descriptor of size d and $S_K = \{(\pi_k, \mu_k, \sigma_k) : k = 1 \ldots K\}$ a set of K semantic sub-groups learnt by the GMM. The FV representation for $\mathbf{X}$ is defined as the conditional probability:

$$FV_X = P(X|S_K) \tag{1}$$
$$= P(X|\mu_k, \sigma_k) \tag{2}$$

By applying Bayesian properties, we can rewrite previous equation as:

$$\phi_k^1 = \frac{1}{N\sqrt{\pi_k}} \sum_{i=1}^{d} \alpha_i(k) \left(\frac{X_i - \mu_k}{\sigma_k} \right), \tag{3}$$

$$\phi_k^2 = \frac{1}{N\sqrt{2\pi_k}} \sum_{i=1}^{d} \alpha_i(k) \left(\frac{(X_i - \mu_k)^2}{\sigma_k^2} - 1 \right), \tag{4}$$

where $\alpha_i(k)$ is the soft assignment weight or the posterior probability of the i-th feature $\mathbf{X}_i$ to the k-th Gaussian [39]. Therefore, the FV representation that defines a fingerprint image is finally obtained by stacking the differences: $\phi = \left[\phi_1^1, \phi_1^2, \ldots, \phi_K^1, \phi_K^2\right]$, thereby resulting a $2 \cdot d \cdot K = 2 \cdot 64 \cdot K$ size vector.

3) VECTOR LOCALLY AGGREGATED DESCRIPTORS (VLAD)

In order to reduce the high-dimension image representation proposed by the FV and BoW approaches, gaining in efficiency and memory usage, we have finally studied the Vector Locally Aggregated Descriptors (VLAD) methodology [33] (see Fig. 1, third row). This is a simplified non-probabilistic version of FV, which models the data distribution from the accumulative distances between a local descriptor $\mathbf{X}$ and its closest visual word $\mathbf{c}$ in the visual vocabulary. Therefore, as in the BoW approach, a visual vocabulary needs to be computed in the first step with the k-means algorithm.

More specifically, a d-dimensional local feature descriptor $\mathbf{X}$ (i.e., dense-SIFT descriptor) can be represented by a VLAD descriptor $\mathbf{V_X}$ of size Kd as follows:

$$\mathbf{V_X} = \sum_{j=1}^{d} \left(\sum_{X:NN(X)=c_i} X_j - c_{i,j} \right), \tag{5}$$

where $\mathbf{X}_j$ and $c_{i,j}$ denote the j-th component of $\mathbf{X}$, and its corresponding closest visual word $\mathbf{c_i}$. In our method, $\mathbf{V_X}$ is subsequently L_2-normalised in order to further improve the classification accuracy.

Finally, it is important to highlight that VLAD also uses PCA for decorrelating training data.

C. CLASSIFICATION

In order to classify the final encoded representations, separated linear SVMs have been used for each encoding approach. SVMs are popular since they perform well in high-dimensional spaces, avoid over-fitting and have good generalisation capabilities. According to [40], when the feature's dimensionality is so greater than the number of instances employed for training, a non-linear mapping does not improve the performance. Therefore, the use of a linear kernel would be good enough to achieve a high classification accuracy.

In order to find the optimal hyperplane separating the bona fide from the attack presentations, the optimisation algorithm bounds the loss from below. Therefore, we have trained a linear SVM as follows: The SVM labels the bona fide samples as +1 and the presentation attacks as -1, thereby yielding the corresponding W (weights) and $\mathbf{b}$ (bias) classifier parameters.

Subsequently, given a feature descriptor $\mathbf{x}$ which was previously yielded by a particular encoding approach (i.e., BoW, FV, VLAD), the final score $s_\mathbf{x}$, which estimates the class of the sample at hand, is computed as the confidence of such

TABLE 2. PAI fabrication materials used in each dataset of the LivDet 2011 - 2019 databases, where U denotes unknown material in the test set.

DB	Dataset	Gelatine	Latex	PlayDoh	Silicone	Silgum	Wood glue	Ecoflex	Body Double	Modasil	Liquid ecoflex	RTV	OOMOO	Gelatine 2	Mix 1	Mix 2
2011	Biometrika	✓	✓			✓	✓	✓								
	Digital P.	✓	✓	✓	✓		✓									
	Italdata	✓	✓			✓	✓	✓								
	Sagem	✓	✓	✓	✓		✓									
2013	Biometrika	✓	✓				✓	✓		✓						
	Italdata	✓	✓				✓	✓		✓						
2015	GreenBit	✓	✓				✓	✓			✓(U)	✓(U)				
	Digital P.	✓	✓				✓	✓			✓(U)	✓(U)				
	Hi_Scan	✓	✓				✓	✓			✓(U)	✓(U)				
	Crossmatch			✓				✓	✓				✓(U)	✓(U)		
2017	GreenBit	✓(U)	✓(U)				✓	✓	✓		✓(U)					
	Digital P.	✓(U)	✓(U)				✓	✓	✓		✓(U)					
	Orcanthus	✓(U)	✓(U)				✓	✓	✓		✓(U)					
2019	GreenBit	✓	✓				✓	✓	✓		✓(U)				✓(U)	✓(U)
	Digital P.	✓	✓				✓	✓	✓		✓(U)				✓(U)	✓(U)
	Orcanthus	✓	✓				✓	✓	✓		✓(U)				✓(U)	✓(U)

decision (i.e., the absolute value of the score is the distance to the hyperplane):

$$s_{\mathbf{x}} = W \cdot \mathbf{x} + \mathbf{b} \tag{6}$$

IV. EXPERIMENTAL EVALUATION

In this section, we evaluate and benchmark the detection performance of each fingerprint encoding scheme described in Sect. III. Specifically, three goals were taken into account for the experimental protocol design: *i*) analyse the impact of the key parameter K (vocabulary size) on the detection performance of the three proposed PAD schemes, *ii*) benchmark the detection performance of our proposals against the top state-of-the-art approaches, and *iii*) study the computational performance of the three fingerprint encoding schemes.

A. EXPERIMENTAL PROTOCOL

The proposed PAD methods were implemented in C++ using the open-source VLFeat library.[2] All the experiments were conducted on an Intel(R) Xeon(R) CPU E5-2670 v2 processor at 2.50 GHz, 378GB RAM.

1) DATABASES

The experiments were conducted on the well-established benchmarks from LivDet 2011 [41], LivDet 2013 [42], LivDet 2015 [43], LivDet 2017 [31], and LivDet 2019 [44]. A summary of the PAI fabrication materials is included in Tab. 2. It is worth noting that we leave out from our experimental evaluation two datasets in LivDet 2013 (i.e. Crossmatch and Swipe) since some anomalies were detected in their fabrication [31].

2) EVALUATION PROTOCOL AND METRICS

To reach the aforementioned objectives, the experimental evaluation considers three different scenarios: *i*) known-material and known capture device, *ii*) known capture device and unknown-material, and *iii*) unknown capture device and cross-database.

[2]http://www.vlfeat.org/

The detection performance is evaluated in compliance with the ISO/IEC 30107-3 [5]: we report the Attack Presentation Classification Error Rate (APCER), which refers to the percentage of misclassified presentation attacks for a fixed threshold, and the Bona Fide Presentation Classification Error Rate (BPCER), which indicates the percentage of misclassified bona fide presentations. We also include the Detection Error Trade-Off (DET) curves of both error rates, as well as the BPCER for a fixed APCER of 10% (BPCER10), 5% (BPCER20) and 1% (BPCER100).

Then, in order to establish a fair benchmark with the existing literature, we report the ACER as the average of the APCER and the BPCER for a fixed detection threshold δ.

B. EXPERIMENTAL RESULTS

1) KNOWN-MATERIAL AND KNOWN CAPTURE DEVICE SCENARIO

EFFECT OF THE SEMANTIC SUB-GROUPS

First, we optimise the algorithms' detection performance in terms of the main key parameter: the visual vocabulary size K. To that end, we focus on the known scenario, in order to avoid a bias due to other variables. We test the following range of values: $K = \{256, 512, 1024, 2048\}$, since $K > 2048$ would yield too long feature vectors, not usable for real-time applications. Tab. 3 reports the ACER values for the adopted K configurations. As it can be observed, the best K values on average are $K = 512$ for FV and $K = 1024$ for VLAD and BoW. Specifically, the FV representation reports an ACER of 2.23%, which is approximately two and three times lower than the ones attained by the remaining encodings (4.88% for VLAD and 6.34% for BoW). This observation, in turn, indicates that FV is able to successfully separate a BP from an AP given a reduced number of semantic sub-groups built by GMM, in contrast to the VLAD and BoW.

EFFECT OF THE FINGERPRINT QUALITY

On the other hand, we also observed in Tab. 3 that the best performing representation (i.e., FV) achieves a poor detection performance for two out of four datasets in LivDet 2015. In particular, it attains an ACER of 4.30% and 6.20% for

TABLE 3. Detection performance, in terms of ACER(%), of our proposed encodings for different K values. The best results per encoding and capture device are highlighted in bold.

DB	Dataset	FV				VLAD				BoW			
		256	512	1024	2048	256	512	1024	2048	256	512	1024	2048
2011	Biometrika	**2.80**	4.10	5.70	6.10	8.40	8.30	8.30	**8.20**	8.10	7.10	**6.40**	7.30
	Digital P.	0.70	**0.30**	**0.30**	**0.30**	2.00	1.30	0.95	**0.60**	2.20	1.40	**1.30**	1.4
	Italdata	3.20	**2.40**	4.50	5.30	**9.70**	16.10	13.3	12.7	16.20	**7.50**	12.70	19.70
	Sagem	1.72	1.60	**1.42**	1.47	3.00	2.65	2.65	**2.36**	6.48	6.53	5.26	**5.21**
2013	Biometrika	**0.30**	0.50	0.50	0.40	2.50	3.10	**1.80**	2.40	3.10	2.50	**1.80**	**1.80**
	Italdata	**0.30**	**0.30**	**0.30**	**0.30**	1.00	0.80	0.80	**0.70**	4.40	3.90	3.70	**2.60**
2015	GreenBit	1.60	**1.30**	1.40	1.80	4.80	**3.60**	3.80	4.50	4.20	**4.00**	4.40	5.00
	Digital P.	7.30	6.50	6.50	**6.20**	9.70	9.40	**8.90**	9.50	16.00	14.70	13.40	**13.00**
	Hi_Scan	4.60	**4.30**	4.50	4.8	6.50	6.80	**5.80**	6.00	11.40	10.50	9.00	**8.30**
	Crossmatch	1.06	**1.03**	**1.03**	**1.03**	3.62	3.62	**2.50**	3.28	7.03	6.21	5.40	**5.34**
	Avg.	2.36	**2.23**	2.62	2.77	5.12	5.57	**4.88**	5.02	7.91	6.43	**6.34**	6.97

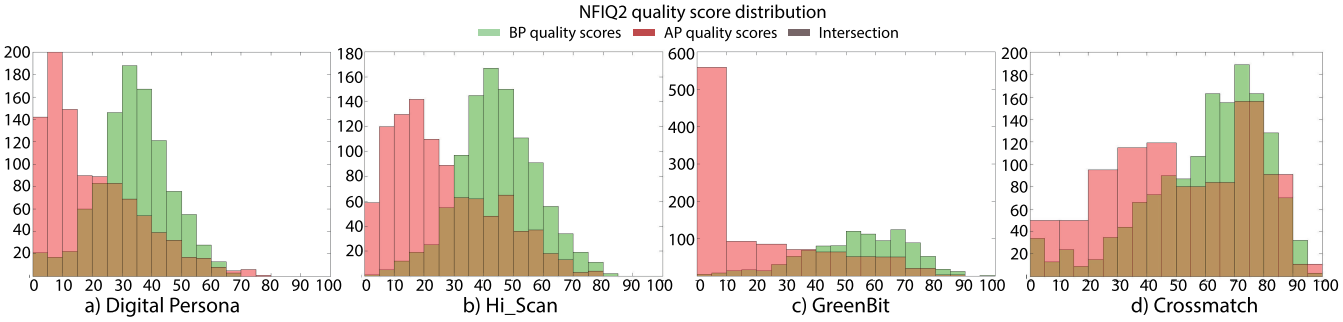


FIGURE 5. NFIQ2 quality distribution for the LivDet 2015 datasets.

Hi_Scan and Digital Persona, which are respectively three and five times worse than the ones reported by GreenBit and Crossmatch. According to [45], most PAD techniques submitted to LivDet 2015 did not perform well due to the small image size. However, by analysing the fingerprint quality provided by the NFIQ2 [46] approach for the entire LivDet 2015 datasets in Fig. 5, we found that most BP images in the Digital Persona and Hi_Scan datasets report a poor NFIQ2 quality, in contrast to the ones in GreenBit and Crossmatch. Whereas 8% and 30% of the fingerprints in Digital Persona and Hi_Scan present a good NFIQ2 quality greater than 50% (good quality), most BP samples in GreenBit (i.e., 63%) and Crossmatch (i.e., 72%) pose a good NFIQ2 quality. Therefore, both capture devices include some sensor technology which produces a high noise degree on the fingerprint samples, and hence also affects the detection performance of most state-of-the-art PAD methods [10], [45], even our approach. This observation is also confirmed in Fig. 6, which reports the detection performance of our best encoding for different fingerprint image quality ranges over the LivDet 2015. This way, we do claim that FV performs better as the fingerprint image quality of BP samples improves.

This handicap is depicted in Fig. 7: a poor fingerprint image quality sample, taken from Digital Persona, is misclassified by our best approach. It is worth noting that one of the key hypothesis of our work is that the fabrication of

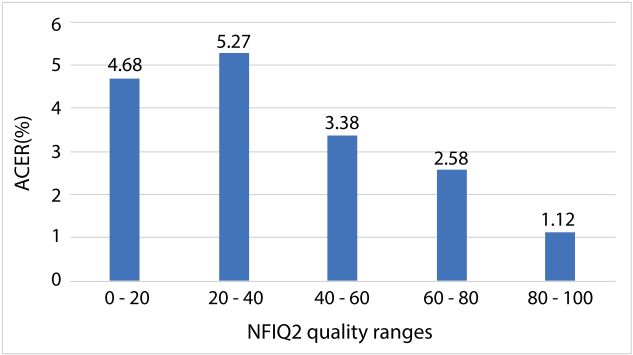


FIGURE 6. ACER evaluation of the FV encoding for several NFIQ2 quality ranges.

PAIs produces artefacts on the ridge pattern which could be correctly detected through the orientation histograms. Therefore, we think that for those capture devices (e.g., Digital Persona and Hi_Scan) which include a high noise degree on the BP ridge pattern, our fingerprint representation-based method is unable to successfully spot an attack presentation. We should also not forget that, those high-noise fingerprint images would not be suitable for a real fingerprint recognition pipeline. Finally, we also do confirm that the orientation field, representing a fingerprint ridge pattern, can be used as a discriminative feature to detect attack presentation attempts

TABLE 4. Benchmark in terms of the ACER(%) with the top state-of-the-art. The best results are highlighted in bold.

DB	Dataset	VGG [10]	TripleNet [11]	FSB-v1 [12]	TinyFCN [18]	FSB-v2 [13]	FLDNet [23]	FV	FUSION
2011	Biometrika	5.20	5.15	2.60	**1.10**	1.24	-	2.80	2.40 ($\alpha = 0.7$, $\beta = 0.0$)
	Digital P.	3.20	1.85	2.70	1.10	1.61	-	0.30	**0.10** ($\alpha = 0.8$, $\beta = 0.0$)
	Italdata	8.00	5.10	3.25	4.75	2.45	-	2.40	**2.20** ($\alpha = 0.8$, $\beta = 0.0$)
	Sagem	1.70	1.23	1.80	1.56	1.39	-	1.42	**1.13** ($\alpha = 0.8$, $\beta = 0.2$)
	Avg.	4.52	3.33	2.59	3.12	1.67	-	1.73	**1.46**
2013	Biometrika	1.80	0.65	0.60	0.35	**0.20**	0.36	0.30	0.30 ($\alpha = 0.9$, $\beta = 0.0$)
	Italdata	0.40	0.50	0.40	0.40	**0.30**	1.35	**0.30**	**0.30** ($\alpha = 0.1$, $\beta = 0.0$)
	Avg.	1.10	0.58	0.50	0.38	**0.25**	0.86	0.30	0.30
2015	GreenBit	4.60	-	2.00	**0.20**	0.68	0.53	1.30	1.30 ($\alpha = 1.0$, $\beta = 0.0$)
	Digital P.	5.64	-	1.76	3.40	**1.12**	3.61	6.20	6.20 ($\alpha = 1.0$, $\beta = 0.0$)
	Hi_Scan	6.28	-	1.08	**0.35**	1.48	2.95	4.30	4.30 ($\alpha = 1.0$, $\beta = 0.0$)
	Crossmatch	1.90	-	0.81	1.09	**0.64**	1.78	1.03	1.03 ($\alpha = 1.0$, $\beta = 0.0$)
	Avg.	4.61	-	1.39	1.26	**0.97**	2.22	3.20	3.20

a) Bona fide presentation NFIQ2 = 29

b) Attack presentation NFIQ2 = 29

FIGURE 7. BP and AP samples which report the same NFIQ2 quality. a) a misclassified BP sample whose ridges include a high noise degree, and b) an AP image with a high noise degree.

whose capture devices do not include a high noise degree over the BP ridge pattern.

STATE OF THE ART BENCHMARK

Once determined the best performing K values, we benchmark, in Tab. 4, our best encoding against the state-of-the-art in terms of the ACER for the known scenario. The lowest value on each row is highlighted in bold. Given that the use of complementary information could improve the recognition capabilities of an approach, we also evaluate the fusion between the three proposed representations (i.e., FV, VLAD, and BoW) using a weighted sum method as follow:

$$s_f = \alpha \cdot s_1 + \beta \cdot s_2 + (1 - \alpha - \beta) \cdot s_3, \tag{7}$$

where $\alpha + \beta \leq 1$, and s_1, s_2 and s_3 represent the individual scores produced by our three encodings. Taking into account that LivDet databases do not include a validation set, the α and β weighted values are computed from each LivDet's training set. In our work, we also experimented the fusion at feature's level between the three fingerprint representations. However, the new large-dimensional vectors together with the poor detection capability reported by VLAD and BoW features led to a detection performance degradation of this fusion. The large feature dimensionality in conjunction with the low number of training samples in datasets do not allow that the SVM finds the entire set of optimal hyperplanes to successfully separate a BP from an AP either.

Taking a look at Tab. 4, we can observe that our FV representation and its fusion with the other encodings achieve the state of the art in most datasets. Specifically, the former reports an average ACER of 1.73% and 0.30% for LivDet 2011 and LivDet 2013, respectively, which outperform four out of five top state-of-the-art approaches, thereby confirming its detection capability for this baseline scenario. In addition, its detection performance is slightly enhanced by the fusion for one out of three databases (i.e., 1.46% vs. 1.73% for LivDet 2011), hence showing that the joint information between a hard and soft assignment leads to a classification improvement. Finally, given that the fingerprint image quality significantly affects the PAD performance of the proposed encodings, the benchmarking methods are unable to outperform on average most state-of-the-art techniques. However, ACERs of 1.30% and 1.03%, which achieve the current state-of-the-art PAD schemes, are respectively reported for the two good-quality fingerprint datasets on the LivDet 2015 (i.e., GreenBit and Crossmatch).

IN DEPTH DETECTION PERFORMANCE ANALYSIS

It should be noted that the main goal of the present work is not only to achieve the best performance at a single operating point (i.e., the ACER is measured for $\delta = 0.5$) but overall for different applications requiring either a low BPCER (i.e., high convenience) or low APCER (i.e., high security), and also under more challenging and realistic conditions (i.e., unknown capture devices or PAI species). Fig. 8 shows the DET curves for the proposed fusion representation over all capture devices employed for the known scenario. As it can be observed, a significant performance can be deployed by the fusion for high security thresholds (i.e., APCER = 1.0%): a remarkable average BPCER100 = 1.98% (vs. 9.68% in FSB-v1 [12], 4.05% in FSB-v2 [13]) for LivDet 2011 and 0.10% (vs. 0.20% in FSB-v1 [12], 0.05% in FSB-v2 [13]) for LivDet 2013. More in detail, for the entire set of capture

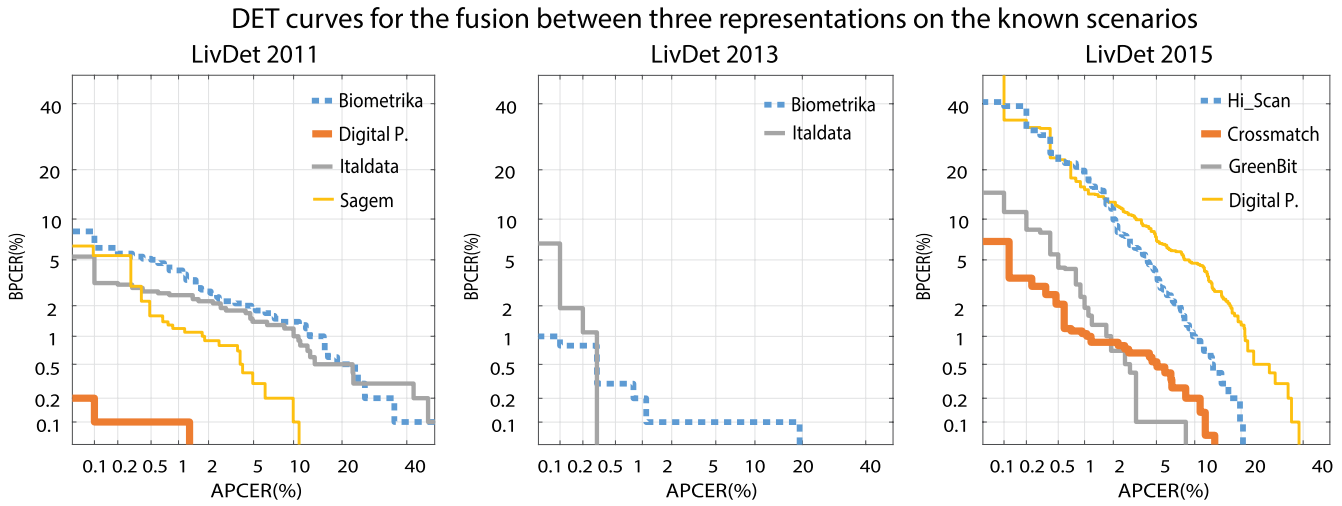


FIGURE 8. Performance evaluation for the fusion between three representations under the known-material and known capture device scenario.

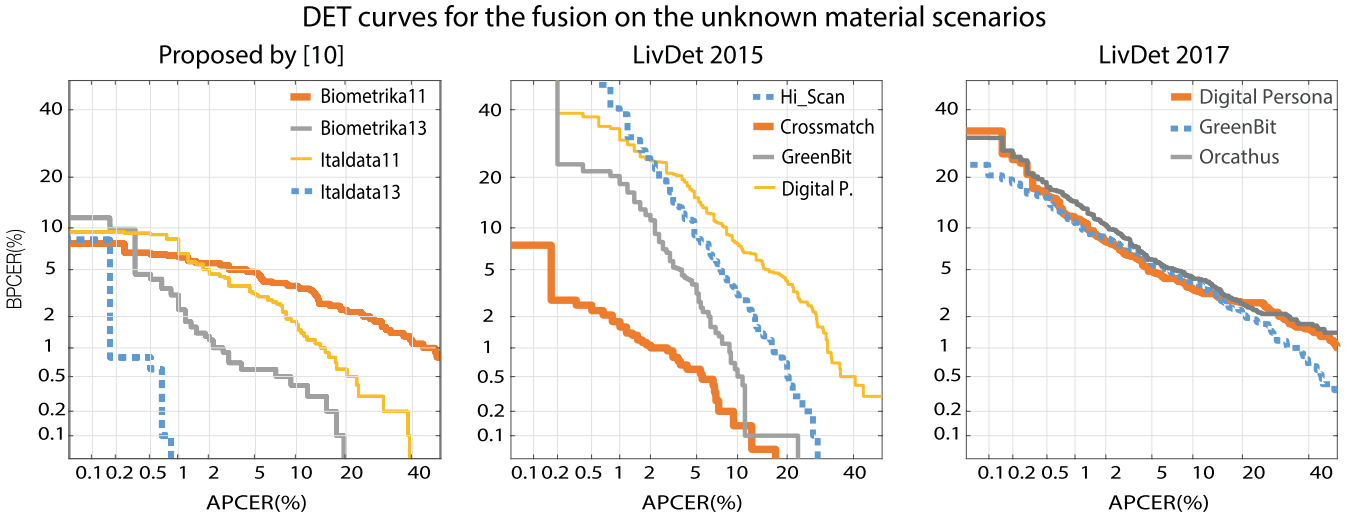


FIGURE 9. Performance evaluation over the unknown PAI species scenarios.

devices in LivDet 2011, a BPCER in range of 0.10% - 8.20% for any APCER $\leq$ 1.0% is reported. Similar results can be consequently observed for all capture devices in LivDet 2013: a BPCER in range of 0.10% - 6.7% for any APCER $\leq$ 1.0%. In contrast, our fusion approach suffers a detection performance decrease for two out of four dataset in LivDet 2015: a joint BPCER100 = 17% for Hi_Scan and Digital Persona, which is 12 times greater than the one reported by GreenBit and Crossmatch (i.e., BPCER100 = 1.45%), confirms the impact of fingerprint image quality on the encoding detection performance.

2) KNOWN CAPTURE DEVICE AND UNKNOWN-MATERIAL SCENARIO

As it was aforementioned, the main goal of our work is faced scenarios with unknown factors. Therefore, we analyse in detail the detection performance of the three fingerprint representations and their fusion for unknown PAI species. This is, both training and test samples were acquired by the same capture device, and species in the test set are unknown for training. For the latter, we select the fixed K, α, and β values obtained for the known-scenarios (see values in Tab. 3). Given that LivDet 2017 database aimed the PAD evaluation for an unknown PAI species [31], we optimised the encodings for that database following the aforementioned protocol in Sect. IV-B1. The performance evaluation for the best performing representation together with the fusion are presented in Tab. 5, and the corresponding DET curves are shown in Fig. 9.

As it can be observed in Tab. 5, the fusion approach slightly outperforms the single FV representation, thereby resulting an average ACER of 2.60%, 5.26% and 5.15% for the adopted unknown PAI species protocols. These results, in turn, achieve the state-of-the-art techniques, hence confirming its

TABLE 5. Detection performance of our encoding methods, in terms of ACER(%), for several **unknown-material** scenarios.

Protocol	Dataset	PAI species Train	PAI species Test	FV	FUSION	FSB-v2 [13]	FLDNet [23]	LivDet 2017 winner [31]†	FSG [17]×
Proposed by [10]	Bio11	EcoFlex, Gelatine, Latex	Silgum, Woodglue	6.33	4.78	**4.60**	-	-	-
	Bio13	Modasil, Woodglue	EcoFlex, Gelatine, Latex	1.00	1.50	1.30	**0.87**	-	-
	Ita11	EcoFlex, Gelatine, Latex	Silgum, Woodglue, Other	3.78	**3.60**	5.20	-	-	-
	Ita13	Modasil, Woodglue	EcoFlex, Gelatine, Latex	**0.30**	0.50	0.60	0.94	-	-
		Avg.		2.85	**2.60**	2.93	-	-	-
LivDet2015	Crossmatch	Body Double, EcoFlex,PlayDoh	Gelatine, OOMOO	**1.34**	**1.34**	-	2.66	-	-
	Digital	EcoFlex_00-50,Latex, Gelatine, Woodglue	Liquid EcoFlex, RTV	8.85	8.85	-	**3.06**	-	-
	Greenbit			4.20	4.20	-	**0.46**	-	-
	Hi_Scan			6.65	6.65	-	**3.38**	-	-
		Avg.		5.26	5.26	-	**2.39**	-	-
LivDet2017⋆	Digital+ Greenbit+ Orcanthus◇	Woodglue, EcoFlex, Body Double	Gelatine, Latex, Liquid EcoFlex	4.92	4.84	-	-	4.41	4.80
				5.46	5.35	-	-	3.56	**2.58**
				5.62	5.62	-	-	6.29	**4.99**
		Avg.		5.33	5.15	-	-	4.75	**4.12**
LivDet2019‡	Digital	Woodglue, EcoFlex, Body Double, latex, Gelatine	Mix 1, Mix 2, Liquid EcoFlex	**6.37**	-	-	-	11.14	16.36
	Greenbit			2.32	-	-	-	0.80	**0.27**
	Orcanthus			2.79	-	-	-	2.55	**2.50**
		Avg.		**3.83**	-	-	-	4.83	6.38

† The overall classification errors reported by the LivDet 2017 winner in this work are the complement of the overall accuracy achieved in [31]
× The overall classification errors reported by FSG in this work are the complement of the overall accuracy achieved in [17]
‡ The overall classification errors reported are the complement of the overall accuracy achieved in [44]
⋆ The ACER results for the encoding fusion was attained at K = 2048 for FV and BoW, and K = 1024 for VLAD.
+ α = 0.9 and β = 0.1.
◇ α = 1.0 and β = 0.0.

soundness for this challenging scenario. It should be noted that for the unknown PAI species protocol described in [10], the FV method shows an ACER below 1.00% for most datasets, with the exception of Biometrika 2011 and Italdata 2011, which contain unknown PAIs fabricated with Silgum. In a previous work [47], we showed how PAIs created with Silgum correctly copied their corresponding fingerprint ridge pattern, thereby making hard to detect by our best fingerprint representation. On the other hand, the ACERs showed by FV and the fusion method for LivDet 2015 also appear to be affected by the aforementioned fingerprint quality. Specifically, Digital Persona and Hi_Scan yield an ACER of 8.85% and 6.65%, respectively, which are worse than the ones attained for the remaining datasets. This fingerprint quality deterioration can be also noted for Digital Persona in LivDet 2017 and 2019: ACERs in range of 4.80% - 16.36% confirm the state-of-the-art shortcomings to identify AP attempts stemming from a high-noise capture device. These error rates also state that those algorithms depend on a careful selection of known PAI species in order to achieve a reliable detection performance over unknown PAI species, in contrast to our proposed representation. This also led that the FV approach outperformed the rest of PAD techniques in the last LivDet 2019 edition, thereby resulting in an overall classification error rate of 3.83%.

Finally, a similarity between the ACERs reported in Tab. 5 and detection performance shown in Fig. 9 can be perceived. Whereas the DET curves for the unknown PAI species protocol described in [10] show an average BPCER100 of 4.53% (vs. 4.90% in FSB-v1 [12], 4.24% in FSB-v2 [13]), the curves for LivDet 2015 and LivDet 2017 report an average BPCER100 of 9.23% and 12.29%, respectively.

3) UNKNOWN CAPTURE DEVICE AND CROSS-DATABASE SCENARIOS

In the last experimental evaluation, we assess the soundness of our proposals in scenarios where different (i.e., unknown) capture devices are used following the unknown capture device and cross-database scenarios proposed by [10].

In the first set of experiments, training and test samples are acquired using different capture devices (i.e., capture device inter-operability analysis). Tab. 6a benchmarks in terms of the ACER the proposed fingerprint representations against the top state-of-the-art PAD techniques. In general and regardless of the particular train-test combination, FV encoding and the fusion method is able to outperform both the other two encoding approaches and the results reported in the literature (i.e., average ACER = 4.08% for FV and average ACER = 5.35% for FV vs. 14.59% for FSB-v2 [13]), which implies a relative improvement of up to 72%.

Consequently, Fig. 10a) shows the corresponding security evaluation for the fusion approach. As it may be observed, training over the Italdata subset yields a better performance at all operating points than training over Biometrika (grey vs. orange and blue vs. yellow curves). Moreover, for a fixed APCER of 1%, the fusion achieves a BPCER100 of 16.24%, which reduces by 75% the top state-of-the-art result (BPCER100 = 52.52% for FSB-v2 [13]), thereby confirming its soundness for this challenging scenario.

In the second experiment, the performance is evaluated over the change of data collection over the same capture device (i.e., train and test over the same capture device, but acquired for LivDet 2011 and LivDet 2013, respectively). We refer to this protocol as cross-database scenario. In Tab. 6b, similar results like the ones reported in Tab. 6a can be observed. Specifically, the FV encoding and

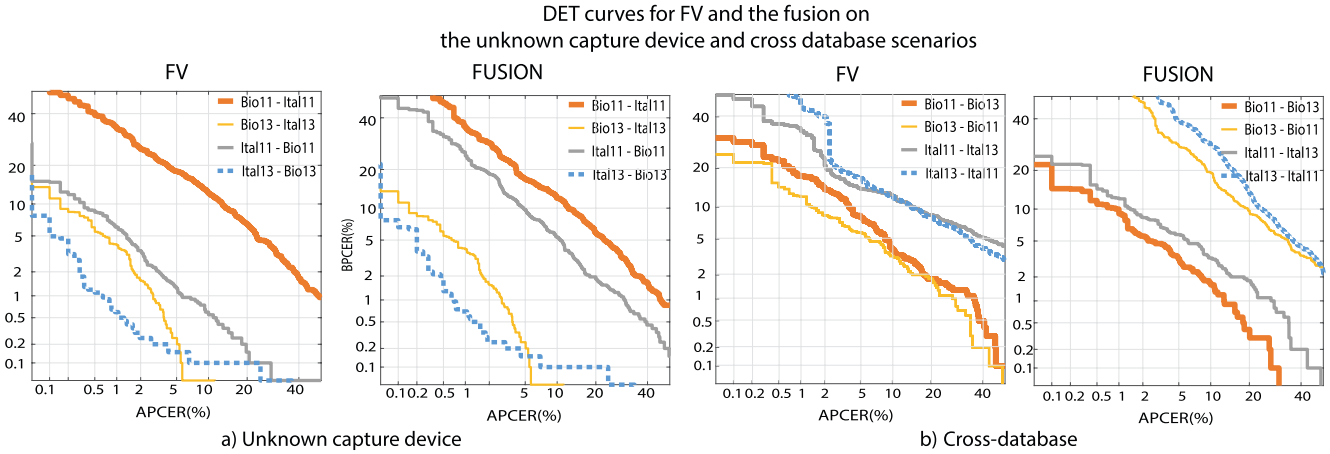


FIGURE 10. **Performance evaluation over the unknown capture device and cross-database scenarios [10].**

TABLE 6. **ACER evaluated on the unknown capture device and cross-database scenarios proposed by [10].**

(a) Unknown capture device protocol.

Train - Test	FV	VLAD	BoW	FUSION	FSB-v2 [13]	FLDNet [23]
Bio11 - Ital11	**11.30**	19.30	20.45	11.45	25.35	-
Bio13 - Ital13	**1.80**	3.50	3.75	**1.80**	4.30	2.10
Ital11 - Bio11	**2.40**	15.20	26.75	7.40	25.21	-
Ital13 - Bio13	**0.80**	1.90	3.00	0.75	3.50	2.90
Avg.	**4.08**	9.98	13.49	5.35	14.59	-

(b) Cross-database protocol.

Train - Test	FV	VLAD	BoW	FUSION	FSB-v2 [13]	TripleNet [11]
Bio11 - Bio13	6.80	15.70	17.80	**4.00**	7.60	14.00
Bio13 - Bio11	12.70	**10.60**	31.20	13.60	31.16	34.05
Ital11 - Ital13	**5.60**	10.00	38.5	**5.60**	6.70	8.30
Ital13 - Ital11	**11.50**	18.10	44.80	17.50	26.16	44.65
Avg.	**9.15**	13.60	33.08	10.18	17.91	25.25

fusion method outperform the remaining PAD techniques (i.e., ACER = 17.91% for FSB-v2 [13] and ACER = 25.25% for TripleNet [11]) by up to a 64% relative improvement.

Regarding to the operational evaluation, we can note, in Fig. 10b), that fingerprint samples acquired by the same capture device at different years produce similar DET curves for FV, in contrast to the fusion, which is affected by error rates produced by BoW. In particular, Biometrika 2011 and Biometrika 2013 yield, for FV, a joint BPCER10 = 3.60%, BPCER20 = 7.10%, and BPCER100 = 15.10% with a standard deviation in range of 0.42% - 3.82%, thereby confirming its soundness in this challenge scenario: current state-of-the-art PAD techniques report a high BPCER100 of 65.06%. Consequently, Italdata 2011 and Italdata 2013 attain a joint BPCER10 = 12.55 ± 0.21, BPCER20 = 15.70% ± 1.98, and BPCER100 = 41.95 ± 9.40.

Finally, a t-SNE visualisation in Fig. 11 for the unknown capture device and cross-database scenarios shows the

TABLE 7. **Computational performance in seconds on LivDet 2015.**

	$K = 256$	$K = 512$	$K = 1024$	$K = 2048$
BoW	0.37	0.38	0.38	0.39
VLAD	1.24	1.33	1.58	1.98
FV	1.17	1.48	2.11	3.39

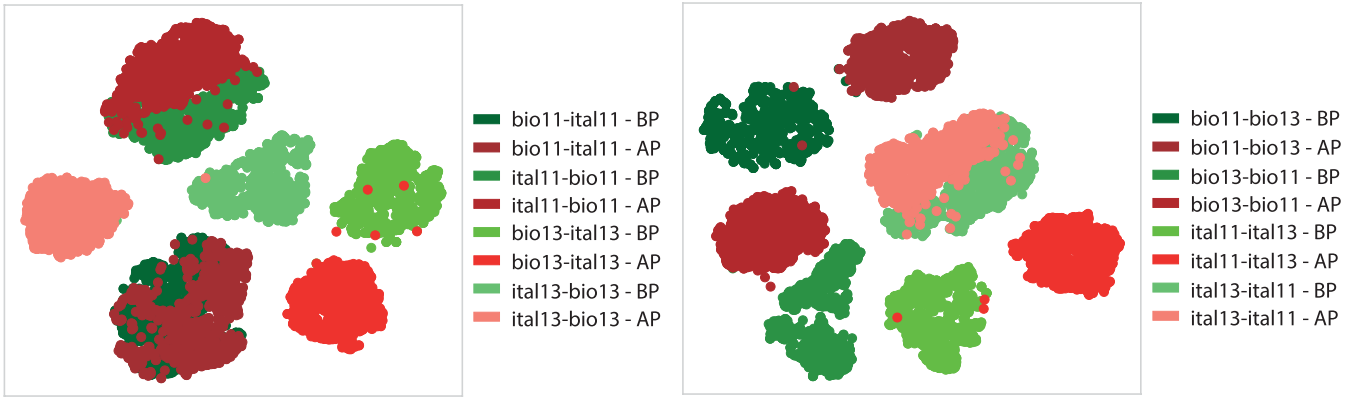
capability of the FV representation to separate an AP from a BP. We can observe that feature spaces for AP samples appear to be, at most cases, closer with each other than with those BP attempts. Even for those testing capture devices such as Biometrika 2011 (see Fig. 11b)), which contains PAI species unknown in the Biometrika 2013 training set, we can note that our approach was able to find a set of semantic sub-groups from known samples to successfully fit those unknown PAI species. This, in turn, confirms the aforementioned hypothesis in Sect. I.

4) COMPUTATIONAL EFFICIENCY

In this last set of experiments, we study the computational efficiency of the proposed fingerprint encodings for different parameter configurations, as reported in Tab. 7. For this purpose, we select the LivDet 2015 database, which contains the largest images. As expected, the time employed by our representation increases as the number of semantic sub-groups grows. Specifically, we found that for the best K value in which the three proposed representations achieved their optimum detection performance, the BoW representation requires 0.38 seconds ($K = 1024$), VLAD 1.58 seconds ($K = 1024$), and FV 1.48 seconds ($K = 512$) to output a decision. The aforementioned detection performance results together with low memory requirements (i.e., approximately 500kb) needed by each trained model lead to a good trade-off between detection accuracy and computational efficiency.

5) FINAL REMARKS

We can summarise the main findings of the experimental evaluation as follows:



a) Unknown capture device b) Cross-database

FIGURE 11. t-SNE visualisation of the FV common feature space for the unknown capture device and cross-database scenarios.

- Among the three encoding proposals (i.e., BoW, FV, and VLAD), the best detection performance is obtained for FV, and the worst for BoW.
- SIFT descriptors representing the fingerprint ridge pattern characteristics such as ridge shape, smoothness and orientation, are suitable to detect attack presentation attempts.
- Experimental evaluation showed that most state-of-the-art approaches including the proposed representations are significantly affected by the fingerprint image quality for those samples acquired with a high noise capture device such as Digital Persona.
- The NFIQ2 quality results showed how the detection performance of the FV encoding improved for better BP fingerprint qualities over the LivDet 2015 database.
- Experimental evaluation also showed that our best fingerprint representation (i.e. FV) is able to outperform, with a high accuracy, the top state-of-the-art in the more realistic and challenges scenarios, in which both known and unknown PAIs are frequently employed.
- Whereas Deep learning-based fingerprint PAD approaches require large databases for optimising thousands of parameters, our proposal attained a high detection performance optimising a small number of them (K) from a small dataset.
- FV encoding can yield an improvement up to 72%, in more complex and realistic scenarios, even for very high security operating points (APCER = 1%).
- Experimental results demonstrated the soundness of our best fingerprint representation to detect AP samples acquired at different years by the same capture device: a BPCER in range of 3.60% - 7.10% with an average standard deviation of 2.03 for higher security thresholds (i.e, 1% $\leq$ APCER $\leq$ 10%).
- Further, a fusion at score's level between three fingerprint representations deployed a performance improvement at most cases, thereby resulting in a BPCER100 in range of 1.98% - 17% in the presence of unknown PAI species. This, in turn, confirmed that the hard quantisation computed by VLAD and BoW can be used as addition information to enhance the soft quantisation built by the FV approach. We think that for those non-improvement cases, a proper tuning of the fusion parameters could enhance its detection performance.
- Fig. 11 showed that unknown PAIs share homogeneous characteristics with each other and heterogeneous with those of bona fide presentations. However, the overlap between some BP and AP samples requires new generative models in order to successfully learn the input data distribution, and hence enhance the PAD generalisation capabilities.

V. CONCLUSION

In this paper, we have proposed a new PAD method based on the combination of local dense-SIFT image descriptors and three different feature encoding approaches (i.e., FV, VLAD, and BoW). The detection performance evaluation conducted over most publicly available LivDet databases showed the soundness of our best fingerprint representation (i.e., FV) in more complex and realistic scenarios where unknown and known attack presentation attempts are carried out. In addition, this best single encoding achieved the highest detection accuracy on the LivDet 2019 competition [29].

In more details, the ISO-compliant evaluation in terms of BPCER and APCER showed one of the main strengths of the FV encoding: the low BPCERs achieved even for very high security operating points (i.e., APCER $\leq$ 1%). Specifically, the FV approach yielded an average BPCER100 of 1.28% for known-scenarios, 8.69% for the unknown material scenarios, and 11% and 24% for the unknown capture device and cross-database scenarios, respectively, thereby achieving the top state-of-the-art results. In addition, a fusion between three encodings through a weighted sum approach showed an improvement of the baselines at most cases, thereby resulting in a BPCER100 in range of 1.98% - 17% in the presence of unknown PAI species.

In summary, previous results indicate that *i*) in the challenging scenarios, orientation histograms computed by the dense-SIFT method correctly represent the ridge pattern, and hence the artefacts produced in the fabrication of PAIs; and *ii*) FV in combination with a generative model as GMM which correctly learned the data distribution dense-SIFT descriptors yielded a new common feature space, which allows successfully detecting both known and unknown PAIs.

Finally, the computational efficiency evaluation showed that BoW encoding attained efficiency results below 400 milliseconds, while VLAD and FV encodings were above 1150 milliseconds. As future work lines, we will improve the computational cost of the FV encodings in order to obtain the best trade-off between detection accuracy and computational efficiency. In order to tackle the fingerprint image quality limitation provided by Digital Persona, we will evaluate other texture descriptors in combination with FV. In addition, we will evaluate a new generative model in order to remove the GMM constraints on the input data distribution.

REFERENCES

[1] D. Maltoni, D. Maio, A. K. Jain, and S. Prabhakar, *Handbook Fingerprint Recognition*. Cham, Switzerland: Springer, 2009.

[2] J. Galbally-Herrero, J. Fierrez-Aguilar, J. Rodriguez-gonzalez, F. Alonso-Fernandez, J. Ortega-Garcia, and M. Tapiador, "On the vulnerability of fingerprint verification systems to fake fingerprints attacks," in *Proc. 40th Annu. Int. Carnahan Conf. Secur. Technol.*, Oct. 2006, pp. 130–136.

[3] J. Galbally, R. Cappelli, A. Lumini, G. Gonzalez-de-Rivera, D. Maltoni, J. Fierrez, J. Ortega-Garcia, and D. Maio, "An evaluation of direct attacks using fake fingers generated from ISO templates," *Pattern Recognit. Lett.*, vol. 31, no. 8, pp. 725–732, Jun. 2010.

[4] T. Matsumoto, H. Matsumoto, K. Yamada, and S. Hoshino, "Impact of artificial 'gummy' fingers on fingerprint systems," in *Proc. Opt. Secur. Counterfeit Deterrence Techn.*, Apr. 2002, pp. 275–289.

[5] *ISO/IEC JTC1 SC37 Biometrics, ISO/IEC 30107-3. Information Technology—Biometric Presentation Attack Detection—Part 3: Testing and Reporting*, Standard ISO/IEC 30107-3, 2017.

[6] A. Zwiesele, A. Munde, C. Busch, and H. Daum, "BioIS study. Comparative study of biometric identification systems," in *Proc. IEEE 34th Annu. Int. Carnahan Conf. Secur. Technol.*, 2000, pp. 60–63.

[7] S. Schuckers, "Presentations and attacks, and spoofs, oh my," *Image Vis. Comput.*, vol. 55, pp. 26–30, Nov. 2016.

[8] A. Rattani, W. J. Scheirer, and A. Ross, "Open set fingerprint spoof detection across novel fabrication materials," *IEEE Trans. Inf. Forensics Security*, vol. 10, no. 11, pp. 2447–2460, Nov. 2015.

[9] Y. Ding and A. Ross, "An ensemble of one-class SVMs for fingerprint spoof detection across different fabrication materials," in *Proc. IEEE Int. Workshop Inf. Forensics Secur. (WIFS)*, Dec. 2016, pp. 1–6.

[10] R. F. Nogueira, R. de Alencar Lotufo, and R. Campos Machado, "Fingerprint liveness detection using convolutional neural networks," *IEEE Trans. Inf. Forensics Security*, vol. 11, no. 6, pp. 1206–1213, Jun. 2016.

[11] F. Pala and B. Bhanu, "Deep triplet embedding representations for liveness detection," in *Deep Learning for Biometrics*. Cham, Switzerland: Springer, 2017, pp. 287–307.

[12] T. Chugh, K. Cao, and A. K. Jain, "Fingerprint spoof detection using minutiae-based local patches," in *Proc. IEEE Int. Joint Conf. Biometrics (IJCB)*, Oct. 2017, pp. 581–589.

[13] T. Chugh, K. Cao, and A. K. Jain, "Fingerprint spoof buster: Use of minutiae-centered patches," *IEEE Trans. Inf. Forensics Security*, vol. 13, no. 9, pp. 2190–2202, Sep. 2018.

[14] T. Chugh and A. K. Jain, "Fingerprint presentation attack detection: Generalization and efficiency," in *Proc. Int. Conf. Biometrics (ICB)*, Jun. 2019, pp. 1–8.

[15] J. J. Engelsma and A. K. Jain, "Generalizing fingerprint spoof detector: Learning a one-class classifier," in *Proc. Int. Conf. Biometrics (ICB)*, Jun. 2019, pp. 1–8.

[16] R. Gajawada, A. Popli, T. Chugh, A. Namboodiri, and A. K. Jain, "Universal material translator: Towards spoof fingerprint generalization," in *Proc. Int. Conf. Biometrics (ICB)*, Jun. 2019, pp. 1–8.

[17] T. Chugh and A. K. Jain, "Fingerprint spoof detector generalization," *IEEE Trans. Inf. Forensics Secur.*, vol. 16, pp. 42–55, 2021.

[18] E. Park, X. Cui, T. H. B. Nguyen, and H. Kim, "Presentation attack detection using a tiny fully convolutional network," *IEEE Trans. Inf. Forensics Security*, vol. 14, no. 11, pp. 3016–3025, Nov. 2019.

[19] E. Marasco and A. Ross, "A survey on antispoofing schemes for fingerprint recognition systems," *ACM Comput. Surv.*, vol. 47, no. 2, p. 28, 2015.

[20] C. Sousedik and C. Busch, "Presentation attack detection methods for fingerprint recognition systems: A survey," *IET Biometrics*, vol. 3, no. 4, pp. 219–233, Dec. 2014.

[21] J. Galbally, S. Marcel, and J. Fierrez, "Biometric antispoofing methods: A survey in face recognition," *IEEE Access*, vol. 2, pp. 1530–1552, 2014.

[22] J. Galbally and M. Gomez-Barrero, "Presentation attack detection in iris recognition," in *Iris and Periocular Biometrics*, C. Busch and C. Rathgeb, Eds. Edison, NJ, USA: IET, Aug. 2017.

[23] Y. Zhang, S. Pan, X. Zhan, Z. Li, M. Gao, and C. Gao, "Fldnet: Light dense cnn for fingerprint liveness detection," *IEEE Access*, vol. 8, pp. 84141–84152, 2020.

[24] D. G. Lowe, "Distinctive image features from scale-invariant keypoints," *Int. J. Comput. Vis.*, vol. 60, no. 2, pp. 91–110, Nov. 2004.

[25] G. Csurka, C. R. Dance, L. Fan, J. Willamowski, and C. Bray, "Visual categorization with bags of keypoints," in *Proc. Int. Workshop Stat. Learn. Comput. Vis. (ECCV)*, 2004, pp. 1–22.

[26] X. Peng, L. Wang, X. Wang, and Y. Qiao, "Bag of visual words and fusion methods for action recognition: Comprehensive study and good practice," *Comput. Vis. Image Understand.*, vol. 150, pp. 109–125, Sep. 2016.

[27] C. H. Lampert, H. Nickisch, and S. Harmeling, "Attribute-based classification for zero-shot visual object categorization," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 36, no. 3, pp. 453–465, Mar. 2014.

[28] J. Sánchez, F. Perronnin, T. Mensink, and J. Verbeek, "Image classification with the Fisher vector: Theory and practice," *Int. J. Comput. Vis.*, vol. 105, no. 3, pp. 222–245, Dec. 2013.

[29] G. Orrá, R. Casula, P. Tuveri, C. Bazzoni, G. Dessalvi, M. Micheletto, L. Ghiani, and G. Luca Marcialis, "LivDet in action–fingerprint liveness detection competition 2019," 2019, *arXiv:1905.00639*. [Online]. Available: http://arxiv.org/abs/1905.00639

[30] R. Tolosana, M. Gomez-Barrero, C. Busch, and J. Ortega-Garcia, "Biometric presentation attack detection: Beyond the visible spectrum," *IEEE Trans. Inf. Forensics Security*, vol. 15, pp. 1261–1275, 2020.

[31] V. Mura, G. Orru, R. Casula, A. Sibiriu, G. Loi, P. Tuveri, L. Ghiani, and G. L. Marcialis, "LivDet 2017 fingerprint liveness detection competition 2017," in *Proc. Int. Conf. Biometrics (ICB)*, Feb. 2018, pp. 297–302.

[32] A. Bosch, A. Zisserman, and X. Munoz, "Image classification using random forests and ferns," in *Proc. IEEE 11th Int. Conf. Comput. Vis.*, Oct. 2007, pp. 1–8.

[33] H. Jegou, F. Perronnin, M. Douze, J. Sanchez, P. Perez, and C. Schmid, "Aggregating local image descriptors into compact codes," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 34, no. 9, pp. 1704–1716, Sep. 2012.

[34] H. Lodhi, C. Saunders, J. Shawe-Taylor, N. Cristianini, and C. Watkins, "Text classification using string kernels," *J. Mach. Learn. Res.*, vol. 2, pp. 419–444, Feb. 2002.

[35] L. J. Gonzalez-Soler, L. Chang, J. Hernandez-Palancar, A. P. Suarez, and M. Gomez-Barrero, "Fingerprint presentation attack detection method based on a bag-of-words approach," in *Proc. Iberoamerican Conf. Pattern Recognit. (CIARP)*, 2017, pp. 263–271.

[36] C. Elkan, "Using the triangle inequality to accelerate k-means," in *Proc. Int. Conf. Mach. Learn. (ICML)*, 2003, pp. 147–153.

[37] A. Vedaldi and A. Zisserman, "Efficient additive kernels via explicit feature maps," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 34, no. 3, pp. 480–492, Mar. 2012.

[38] F. Perronnin, J. Sánchez, and T. Mensink, "Improving the Fisher kernel for large-scale image classification," in *Proc. Eur. Conf. Comput. Vis. (ECCV)*, 2010, pp. 143–156.

[39] K. Simonyan, O. M. Parkhi, A. Vedaldi, and A. Zisserman, "Fisher vector faces in the wild," in *BMVC*, vol. 2, no. 3, 2013, p. 4.

[40] C. W. Hsu, C. C. Chang, and C. J. Lin, "A practical guide to support vector classification," Taipei, Taiwan, Tech. Rep., 2003. [Online]. Available: https://www.csie.ntu.edu.tw/~cjlin/papers/guide/guide.pdf

[41] D. Yambay, L. Ghiani, P. Denti, G. L. Marcialis, F. Roli, and S. Schuckers, "LivDet 2011-fingerprint liveness detection competition 2011," in *Proc. 5th IAPR Int. Conf. Biometrics (ICB)*, Mar. 2012, pp. 208–215.

[42] L. Ghiani, D. Yambay, V. Mura, S. Tocco, G. L. Marcialis, F. Roli, and S. Schuckers, "LivDet 2013 fingerprint liveness detection competition 2013," in *Proc. Int. Conf. Biometrics (ICB)*, Jun. 2013, pp. 1–6.
[43] V. Mura, L. Ghiani, G. L. Marcialis, F. Roli, D. A. Yambay, and S. A. Schuckers, "LivDet 2015 fingerprint liveness detection competition 2015," in *Proc. IEEE 7th Int. Conf. Biometrics Theory, Appl. Syst. (BTAS)*, Sep. 2015, pp. 1–6.
[44] G. Orru, R. Casula, P. Tuveri, C. Bazzoni, G. Dessalvi, M. Micheletto, L. Ghiani, and G. L. Marcialis, "LivDet in Action–Fingerprint liveness detection competition 2019," in *Proc. Int. Conf. Biometrics (ICB)*, Jun. 2019, pp. 1–6.
[45] L. Ghiani, D. A. Yambay, V. Mura, G. L. Marcialis, F. Roli, and S. A. Schuckers, "Review of the fingerprint liveness detection (LivDet) competition series: 2009 to 2015," *Image Vis. Comput.*, vol. 58, pp. 110–128, Feb. 2017.
[46] E. Tabassi. (2015). *Development of NFIQ 2.0*. Accessed: Aug. 24, 2020. [Online]. Available: https://www.nist.gov/services-resources/software/development-nfiq-20
[47] L. J. Gonzalez-Soler, M. Gomez-Barrero, L. Chang, A. P. Suarez, and C. Busch, "On the impact of different fabrication materials on fingerprint presentation attack detection," in *Proc. Int. Conf. Biometrics (ICB)*, Jun. 2019, pp. 1–6.

LÁZARO JANIER GONZÁLEZ-SOLER received the B.Sc. degree in mathematics and computer science from the University of Havana, in 2014. He joined the Advanced Technologies Application Center (CENATAV), Havana, Cuba, to computer science graduate training. He is currently pursuing the Ph.D. degree with the da/sec Group, National Research Center for Applied Cybersecurity (ATHENE), Germany. His principal research interests include biometric and machine learning. specifically, biometric presentation attack detection for fingerprint, iris, and facial characteristics.

MARTA GOMEZ-BARRERO (Member, IEEE) received the M.Sc. degrees in computer science and mathematics and the Ph.D. degree in electrical engineering from the Universidad Autonoma de Madrid, Spain, in 2011 and 2016, respectively. From 2016 to 2020, she was a Postdoctoral Researcher with the National Research Center for Applied Cybersecurity (ATHENE), Hochschule Darmstadt, Germany. She is currently a Professor for IT-Security and technical data privacy at Hochschule Ansbach, Germany. Her current research interests include security and privacy evaluations of biometric systems, presentation attack detection (PAD) methodologies, and biometric template protection (BTP) schemes. She has coauthored more than 70 publications and chaired special sessions and competitions at international conferences.. She received a number of distinctions, including the EAB European Biometric Industry Award, in 2015, the Best Ph.D. Thesis Award by the Universidad Autonoma de Madrid for the term 2015–2016, the Siew-Sngiem Best Paper Award at ICB 2015, the Archimedes Award for young researches from Spanish MECD, and the Best Poster Award at ICB 2013. She is an Associate Editor of the *EURASIP Journal on Information Security* and represents the German Institute for Standardization (DIN) in ISO/IEC SC37 JTC1 SC37 on biometrics

LEONARDO CHANG received the bachelor's degree (Hons.) from CUJAE University, Havana, Cuba, in 2007, and the M.Sc. and Ph.D. degrees in computer science from the National Institute for Astrophysics, Optics, and Electronics of Mexico, in 2010 and 2015, respectively. He was a Researcher with CENATAV, Cuba, from 2007 to 2017. He is currently a full-time Researcher and a Professor with the Tecnologico de Monterrey, Mexico. His research interests include biometrics, object recognition, and video-surveillance applications. He has published several papers in top journals and conferences.

AIREL PÉREZ-SUÁREZ received the B.Sc. degree in computer science from Havana University, in 2002, and the M.Sc. and Ph.D. degrees in computational sciences from the National Institute of Astrophysics, Optics and Electronics (INAOE), in July 2008 and July 2011, respectively. He is currently an Associate Researcher with the Data Mining Department, Advanced Technologies Application Center (CENATAV), Cuba. His research interests include clustering and data mining, such as supervised classification, frequent patterns, association rules, emerging patterns, and community detection on social networks. He has been a member of the Cuban Society of Mathematics and Computation and the Cuban Association for Pattern Recognition, since 2005.

CHRISTOPH BUSCH (Senior Member, IEEE) holds a joint appointment at the Faculty of Computer Science, Hochschule Darmstadt (HDA), Germany. He has been a Lecturer of biometric systems with the Technical University of Denmark (DTU), since 2007. He is currently a member of the Department of Information Security and Communication Technology (IIK), Norwegian University of Science and Technology (NTNU), Norway. He has coauthored more than 400 technical articles and has been a speaker at international conferences. He is a convenor of WG3 in ISO/IEC JTC1 SC37 on biometrics and an active member of CEN TC 224 WG18. He served for various program committees, such as NIST IBPC, ICB, ICHB, BSI-Congress, GI-Congress, DACH, WEDELMUSIC, and EUROGRAPHICS. He is also an Appointed Member of the Editorial Board of the *IET Biometrics* journal and the IEEE TRANSACTIONS ON INFORMATION FORENSICS AND SECURITY. He served as a Reviewer for several conferences, journals, and magazines, such as ACM-SIGGRAPH, ACM-TISSEC, the IEEE COMPUTER GRAPHICS AND APPLICATIONS, the IEEE TRANSACTIONS ON SIGNAL PROCESSING, the IEEE TRANSACTIONS ON INFORMATION FORENSICS AND SECURITY, the IEEE TRANSACTIONS ON PATTERN ANALYSIS AND MACHINE INTELLIGENCE, and the *Computers and Security* journal (Elsevier). On behalf of Fraunhofer, he chairs the biometrics working group of the TeleTrusT Association and the German standardization body on biometrics (DIN-NIA37).