Local feature encoding for unknown presentation attack detection: An analysis of different local feature descriptors

Lázaro J. González-Soler | Marta Gomez-Barrero | Jascha Kolberg | Leonardo Chang | Airel Pérez-Suárez | Christoph Busch

1Biometrics and Internet Security Research Group, Hochschule Darmstadt, Darmstadt, Germany
2Department of Computer Science, Hochschule Ansbach, Ansbach, Germany
3Tecnológico de Monterrey, Campus Estado de México, Estado de México, México
4Advanced Technologies Application Center (CENATAV), La Habana, Cuba

Abstract
In spite of the advantages of using fingerprints for subject authentication, several works have shown that fingerprint recognition systems can be easily circumvented by means of artificial fingerprints or presentation attack instruments (PAIs). In order to address that threat, the existing presentation attack detection (PAD) methods have reported a high detection performance when materials used for the fabrication of PAIs and capture devices are known. However, for more complex and realistic scenarios where one of those factors remains unknown, these PAD methods are unable to correctly separate a PAI from a real fingerprint (i.e. bona fide presentation). In this article, a new PAD approach based on the Fisher Vector technique, which combines local and global information of several local feature descriptors in order to improve the PAD generalisation capabilities, was proposed. The experimental results over unknown scenarios taken from LivDet 2011 to LivDet 2017 show that our proposal reduces the top state-of-the-art average classification error rates by up to four times, thereby making it suitable in real applications demanding high security. In addition, the best single configuration achieved the best results in the LivDet 2019 competition, with an overall accuracy of 96.17%.

1 | INTRODUCTION

Over the last decades, biometric systems have shown to be a reliable and efficient alternative to traditional credential-based access control systems, being thus deployed in a wide range of applications. As a consequence, enhancing the security of biometric systems is of utmost importance. In particular, among the different attacks listed in the literature, the so-called attack presentations (APs) directed to the capture device have received a considerable attention in the recent past [1]. In those attacks, a subject tampers the system by providing a presentation attack instrument (PAI) to the capture device in order to impersonate someone else (i.e. impostor), or to avoid being recognised when he is included in a black-list (i.e. identity concealer).

In order to automatically determine whether a sample stems from a live individual (i.e. it is a bona fide presentation—BP) or from an attack presentation, several presentation attack detection (PAD) approaches have been proposed in the literature [2–5]. Those PAD methods can be broadly divided into hardware or software based. The former detect life signs such as temperature and electric resistance [6], pulse oximetry [7], blood pressure [8], blood flow [9,10], odour [11], or the response within the short-wave infrared spectral band [12,13], by including dedicated hardware. The main drawback of these methods is that they are complex and add additional costs due to the specific hardware needed [14]. Therefore, we focus in the remaining of this article on software-based approaches.

For the particular case of fingerprint recognition systems, the LivDet series competitions [15] have been established as a
common benchmark for software-based PAD techniques, and most works are evaluated on the corresponding freely available databases. In the last editions, deep-learning-based techniques have outperformed most PAD methods built upon handcrafted features [16–18]. In spite of the remarkable detection performance achieved, those approaches still fail to detect unknown PAIs (i.e. not utilised for training the networks or classifiers), or to detect PAIs captured with different capture devices. A reason can be that these methods learn several filters with BP samples, the FV technique aggregates a large set of local descriptors into a high dimensional vector by fitting a parametric generative model such as Gaussian Mixture Model (GMM) [30]. In a nutshell, the FV representation describes how the distribution of a set of local descriptors, extracted from unknown PAIs, differs from the known PAI distribution previously learned by the adopted generative model: this is the so-called ‘probabilistic visual vocabulary’. The FV has shown a high performance in image classification and retrieval tasks [31,32].

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 database acquired with a single Crossmatch capture device in [19]. In particular, they grouped the features extracted by the Fingerprint Spoof Buster PAD method (FSB) [16] 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.

In addition, Gonzalez-Soler et al. proposed in [21] 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, they reported an average classification error rate under 5%. Keeping in mind the desired goal, the authors extended the previous proposal by evaluating the combination between BoW and other two image encodings: Fisher Vector (FV) and Vectors of Locally Aggregated Descriptors (VLAD) [22]. The experimental evaluation on realistic and more challenging unknown scenarios taken from LivDet 2011, 2013, and 2015 databases reported high performance values which substantially outperform the top state of the art. In spite of those achievements, the proposed method in [22] still fails for some types of capture devices including a high noise degree or unwanted noise in the ridge’s pattern of fingerprints (i.e. low quality fingerprint samples).

Inspired by this previous work, we study in this article several well-known texture descriptors in combination with the FV feature encoding for fingerprint PAD purposes. The use of FV encoding instead of BoW or VLAD has led to an increased detection performance, and hence a generalisation capability improvement. We explore the impact of several materials commonly used in detail for the fabrication of the PAIs on the detection performance of several descriptors in combination with FV as follows: a dense version of Scale Invariant Feature Transform (SIFT) [23], Binarised Statistic Invariant Features (BSIF) [24], Local Binary Pattern (LBP) [25], Histogram of Oriented Gradients (HOG) [26], Speed-Up Robust Features (SURF) [27], Binary Robust Independent Elementary Features (BRIEF) [28], and Oriented FAST and Rotated BRIEF (ORB) [29]. All these descriptors have reported remarkable results in several computer vision tasks. By assuming that the unknown PAIs share homogeneous features such as texture, shape, and appearance with known PAIs and heterogeneous characteristics with BP samples, the FV technique aggregates a large set of local descriptors into a high dimensional vector by fitting a parametric generative model such as Gaussian Mixture Model (GMM) [30].

Table 1 reports the best performances achieved by the top state-of-the-art on different unknown attack scenarios.

In a nutshell, the FV representation describes how the distribution of a set of local descriptors, extracted from unknown PAIs, differs from the known PAI distribution previously learned by the adopted generative model: this is the so-called ‘probabilistic visual vocabulary’. The FV has shown a high performance in image classification and retrieval tasks [31,32].

### 1.1 Contributions

In summary, the main contributions of this work with respect to our preliminary study in [22] are as follows:

- New fingerprint representation to accurately identify artefacts produced by materials used in the fabrication of PAIs, which combines the generalisation capability of FV with different well-known descriptors, namely SIFT, SURF, BSIF, LBP, HOG, BRIEF, and ORB.
- NFIQ2.0 evaluation showing that the detection performance of most state-of-the-art techniques is affected by the fingerprint ridge quality.
- A fusion between the three best performing descriptors, which yields, unlike [22], a reliable detection performance for low fingerprint ridge quality datasets.
- In order to evaluate the performance of our proposed FV-based representation and to allow the reproducibility of the presented results, a thorough evaluation compliant with the international standard ISO/IEC 30,107-3 on PAD [33] is conducted over several known and unknown scenarios taken from the LivDet 2011, LivDet 2013, LivDet 2015, and LivDet 2017 benchmarks.

The remainder of this article is organised as follows: in Section 2, related works on the PAD are described. Section 3 presents the proposed encoding method, and Section 3.1 provides a brief description about selected texture descriptors. The experimental protocol carried out in this work is explained in Section 4. The experimental results benchmarking the performance of our proposal against the top state-of-the-art techniques, is described in Section 5. Finally, conclusions and future work directions are included in Section 6.

| TABLE 1  | Summary of the best state-of-the-art results in terms of Average Classification Error Rates (ACER) on challenging scenarios |
|------------|--------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| CNNs       | Cross-materials | Cross-sensor | Cross-database |
| FSBv2 [16] | 2.16%            | 14.59%       | 17.91%          |
| FSBv1 [17] | 2.29%            | 16.60%       | 18.90%          |
| VGG [18]   | 8.19%            | 19.80%       | 30.70%          |
| TripleNet [20] | 5.86%            | 25.25%       | 15.20%          |

Abbreviations: ACER, average classification error rates; FSB, Fingerprint Spoof Buster; VGG, Visual Geometry Group.
2 RELATED WORK

Taking into account that some properties such as morphology, smoothness, and ridge-valley structure may be different between a bona fide presentation (BP) and attack presentation (AP), several texture-based methods have been proposed in the literature [34–37]. However, the aforementioned methods based on handheld features have been significantly outperformed by deep-learning-based approaches. In particular, Nogueira et al. [18] carried out a performance benchmark between three classic Convolutional Neural Networks (CNN): CNN-Random [38], which only uses random filter weights drawn from a Gaussian distribution, CNN-AlexNet [39], pretrained on the ILSVRC 2012, and CNN-Visual Geometry Group (VGG) [40], a 19-layer CNN which achieved the second place in the object detection task of the ImageNet 2014 challenge. Among the previous CNN models, VGG achieved the best results on the LivDet 2015 competition with an overall accuracy of 95.5%. On the other hand, the ACEr reached an 8.0% for the LivDet 2011 database. This is due to the main limitation of this proposal: features are learnt from a whole image with a fixed size of 227×227 pixels. Given that many samples within the LivDet datasets are larger than 640×480 px., and they have an approximate area of the region of interest (ROI) of 210×275 pixels, the ROI of the resized image is not large enough to allow an efficient AP detection.

In contrast to that whole image-based approach, Pala and Bhanu [20] proposed a metric learning method based on a deep triplet convolutional network, which were fed with a single fixed size patch randomly extracted from each training set image. The main limitation of this proposal is that the training patches were randomly extracted from each fingerprint image. Specifically, several patches extracted from Italdata 2011 could be from the blank region of the image thereby resulting in a high ACEr of 5.1%.

Based on the fact that PAIs produce spurious minutiae on a fingerprint image, Chugh et al. [16,17] proposed a deep learning framework for independently classifying local patches around minutiae extracted from a fingerprint image. A fingerprint sample is then classified by computing the average between AP scores of fixed local patches. To that end, the authors used the inception-v2 [17] and mobileNet [16] networks, respectively. In addition, they showed an example in which the independent classification of local patches has the advantages of finding spoof regions inside a PAI image. In this context, ACEr results achieved by this proposal on the LivDet 2011 (1.67%), LivDet 2013 (0.25%) and LivDet 2015 (0.97%) corroborate the soundness of extracting and of classifying local features around spurious minutiae on the PAD task. In spite of the excellent results reported, this approach still shows problems for some unknown attack scenarios.

In a subsequent study, Chugh and Jain [41] analysed different PAI fabrication materials, and from a set of 12, they showed how using only six materials for training is enough to achieve a state-of-the-art detection performance. However, it should be noted that those materials have to be carefully selected in order to cover the widest spectrum of possible PAI species, thereby requiring an ad hoc analysis for each capture device and scenario.

Generally, the main drawback of the aforementioned texture-based methods is that they are based on Supervised Learning and therefore they depend both on the material used for the fabrication of the PAI and the capture device for acquiring the fingerprint images. In this context, results achieved by learnable-based proposals [16–18,42,43] have shown to be effective in a scenario where both capture devices and materials are known a priori. However, for those scenarios where materials and/or recepries, or capture devices are unknown, these PAD methods still report a lower effectiveness. The main reason behind that limitation is that they do not find a common feature space where unknown PAIs are closer to known attack presentations and further apart from the bona fide presentations.

3 PROPOSED APPROACH

Building upon our previous work [22], Figure 1 shows an overview of the proposed PAD approach, which consists of three main steps: (i) local features are extracted from a fingerprint sample, both real- and binary-valued (see Section 3.1); (ii) an unsupervised GMM or Bernoulli Mixture Model (BMM) (i.e. feature space probabilistic model) learns the distribution of the aforementioned decorrelated features, which are subsequently encoded by computing the gradient of the sample log-likelihood with respect to the learned model parameters (see Section 3.2), and (iii) a BP/AP decision is finally carried out by a linear support vector machine, SVM (see Section 3.3).

3.1 Local descriptors

As it can be observed in Figure 1, we analyse several continuous and binary descriptors, which are briefly described in this section. In particular, we have considered (i) gradient- (SIFT, SURF, PHOG), (ii) intensity difference- (BRIEF, ORB), and (iii) texture-based (LBP, BSIF) features. This way, different aspects from the fingerprint samples can be analysed, and eventually fused in order to achieve a more robust PAD scheme. In addition, the reason behind choosing not only continuous but also binary descriptors lays on their higher efficiency at the cost of a small performance loss for other tasks.

In general, it should be noted that all descriptors are computed over the same set of fixed points on a regular grid with a stride $S$ of 3 pixels, and utilising four patches with different window sizes $\sigma$: {4, 6, 8, 10}.

SIFT [23] is one of the most popular histogram-based descriptors due to its robustness to changes in scale, translation, rotation, and other imaging parameters. In our proposal, we follow the SIFT computation in [22] where SIFT are densely extracted at fixed points on a regular grid with an uniform spacing (e.g. 3 pixels). In order to represent sizeable
artefacts produced in the fabrication of PAlS, descriptors are computed over four circular patches with different radii \( \sigma \). Therefore, each point in the grid is represented by four SIFT descriptors. Such variant of the normal SIFT computation is known in the literature as dense SIFT, which is illustrated in Figure 2. To efficiently compute the descriptors, we use the implementation provided in [44], which delivers a speed-up of up to 60x by (i) exploiting the uniform sampling and overlapping between descriptors, and (ii) using linear interpolation with integral image convolution.

SURF [27] is a keypoint-based descriptor, like SIFT, which uses the Haar wavelet transform to approximate the image gradient. In particular, SURF computes the first-order Haar wavelet responses in the \( x \) and \( y \) directions at the orientation assignment step. Similarly to SIFT descriptors, the area around the interest keypoint is subsequently divided into \( 4 \times 4 \) subregions, and the Haar wavelet responses are computed and \( L_2 \) normalised. The final feature vector is the concatenation of the accumulated wavelet responses in each direction and the summation of its absolute values, thus leading to a 64-dimensional vector per keypoint. In our methodology, we selected the 128-dimensional variant, which also includes the first Haar wavelet responses on diagonal directions.

HOG [26] is a local image descriptor capturing the intensity gradients and edge directions to describe the shape and appearance of an object within an image. As the previous descriptors, the HOG features are computed over localised cells. Therefore, it is invariant to geometric and photometric transformations. In this particular case, the cells comprise usually \( 8 \times 8 \) pixels, and a histogram of edges orientation within that cell is computed. Afterwards, cell blocks of \( 16 \times 16 \) pixels are normalised, in order to provide better illumination invariance. In our implementation, we used a multi-scale HOG extension named Pyramid HOG (PHOG), which has reported good results in static facial expression analysis [45] and fingerprint PAD [46]. In this case, the gradient is joined at several pyramid levels, and a histogram is computed for each grid.

LBP [25] is a texture descriptor originally developed for two-dimensional texture analysis, which has obtained excellent results in multiple tasks. It is invariant to rotation, illumination, and orientation changes. More specifically, it represents an image with a histogram of uniform patterns corresponding to
micro-features in the image. These histograms allow capturing both shape and textural features from an image. In our methodology, a multi-resolution analysis is included, by computing the aforementioned histograms on different window sizes. In more details, let \( X \) be a circular image patch with radii \( \sigma \) and \( S \) pixels around the centre. Then, the LBP descriptor is defined as

\[
\text{LBP}_S^{\sigma} = \sum_{i=0}^{\sigma-1} f(g_i - g_c)2^i,
\]

where \( g_i \) with \( i = 0 \ldots S - 1 \) are gray intensity values around the centre \( g_c \) in the image patch. \( f(g_i - g_c) \) is defined as

\[
f(g_i - g_c) = \begin{cases} 
1, & g_i - g_c \geq 0 \\
0, & g_i - g_c < 0 
\end{cases}
\]

In order to capture more information and thereby increase the descriptor distinctiveness, we compute several LBP patterns by combining various radii \( \sigma \). The LBP histograms are subsequently built from those patterns at different scales by varying the window size and sliding over the whole image. Finally, the computationally efficient implementation provided in [47] is used.

BSIF [24] is a local image descriptor computed by binarising the responses of a given image to a set of pre-learned filters to obtain a statistically meaningful representation of the data. More specifically, let \( X \) be an image patch of size \( l \times l \) and \( \mathcal{W} = \{W_1, \ldots, W_N\} \) a set of linear filters of the same size as \( X \). Then, we compute binarised responses \( b_n \):

\[
b_n = \begin{cases} 
1 & \sum_{u,v} W_n(u, v) X(u, v) > 0 \\
0 & \text{otherwise}
\end{cases}
\]

All the filter responses \( b_n \) are subsequently stacked to form a bit string \( b \) with size \( N \) for each pixel. Subsequently, \( b \) is transformed to a decimal value, and then a \( 2^N \) histogram for \( X \) is computed. In our work, 60 filter sets with different sizes \( l = \{3, 5, 7, 9, 11, 13, 15, 17\} \) and number of filters \( N = \{5, 6, 7, 8, 9, 10, 11, 12\} \) were obtained from [24].

Like the SIFT computation, the BSIF histograms are densely extracted over a regular grid with a fixed stride \( S \) of three pixels, and for each point on the grid, histograms are computed over four circular patches \( \sigma \), as depicted in Figure 3b). Therefore, each point in the grid is represented by four BSIF histograms.

Given that the BSIF histograms are extracted for local patches with a small size, they then become sparse vectors as the number of linear filters \( N \) increases. Therefore, we followed the BSIF reduction strategy in [48] and represented each \( 2^N \) BSIF histogram as a 128-component vector by summing the elements for each sequential \( 2^N/128 \) subset in the original histogram, as illustrated in Figure 3c). This representation reduces the storage requirements down to 12.5% for \( N = 10 \) or 3.1% for \( N = 12 \).

**FIGURE 3** Binarised Statistic Invariant Features (BSIF) descriptors computed from \( N = 9 \) filters of size \( l = 5 \). (a) Fingerprint image, (b) BSIF histograms, computed densely at fixed points on a regular grid, with a fixed stride \( S \) for four local patches with different window size, and (c) a reduced BSIF histogram

**BRIEF** [28] is a binary noise-resistant local descriptor, whose computation time is two orders of magnitude faster than SIFT. This is achieved by exploiting the fact that image patches can be efficiently classified on the basis of a relative small number of pairwise intensity comparisons \( r \). Thus, the BRIEF binary descriptor represents a smoothed patch like a bit string constructed from a set of binary intensity tests. More specifically, let \( X \) be a square smoothed image patch, then a binary test \( r \) can be defined as

\[
r(X; x, y) = \begin{cases} 
1 & \text{if } X(x) < X(y) \\
0 & \text{otherwise}
\end{cases}
\]

where \( x \) and \( y \) are locations in \( X \), and \( X(x) \) is the gray value of \( X \) at \( x \). Previous locations are randomly pre-fixed according to a Gaussian distribution around the patch centre. Finally, by using a set of \( \eta \) binary tests, we can obtain a \( \eta \)-bitstring as follows:

\[
f_\eta X = \sum_{i=0}^{\eta-1} r(X; x_i, y_i)2^i.
\]

In our implementation, we select \( \eta = 256 \), since it has shown a better trade-off between effectiveness and efficiency in many real applications [49].

**ORB** [29] is a binary descriptor built upon BRIEF [28] and Features from Accelerated Segment Test (FAST) [50], which additionally provides rotation invariance. The algorithm starts by detecting FAST points in the image, at different scale pyramid levels, and by adding an effective measure of corner orientation, to conform the final FAST keypoint orientation (oFAST) features. Then, a rotation aware BRIEF (rBRIEF)
descriptor is computed and combined with oFAST to obtain the final ORB descriptor.

To elaborate, rBRIEF first steers the BRIEF descriptor according to the orientation of the keypoints, \( \theta \). To that end, rBRIEF discretises \( \theta \) to increments of \( 2\pi/30 \) (12°), and constructs a lookup table of precomputed BRIEF patterns, thereby obtaining rotation invariant features in an efficient manner. However, steering BRIEF leads to a loss of variance in the responses, and thus to less discriminative features. In addition, both BRIEF and its steered version show some correlation in the tests. To tackle these issues, ORB runs a greedy search among all responses, and thus to less discriminative features. In addition, however, steering BRIEF leads to a loss of variance in the responses, which we recover by rotating the descriptor in an efficient manner.

Features are more robust to new samples, which may stem from the fixed descriptors extracted on the previous step. In particular, a GMM on \( \mathbb{R}^d \) is defined as

\[
G_k(x) = \frac{1}{\sqrt{2\pi\sigma_k^2}} \exp \left( -\frac{(x - \mu_k)^2}{2\sigma_k^2} \right)
\]

and means close to 0.5, as well as being uncorrelated. The result is obtained by stacking the differences \( \phi_G = [\phi_1^G, \phi_2^G, \ldots, \phi_K^G] \), thereby resulting in a 2 \( \times \) \( 64 \cdot K \) size vector.

On the other hand, for encoding binary features (Figure 1, bottom) we train a BMM, whose \( K \)-components are represented by the mixture weights (\( \pi_k^B \)) and means (\( \mu_k^B \)), with \( k = 1 \ldots K \). Therefore, a closed-form approximation of FV representation is given by

\[
F_{\phi_{kd}} = \left( \sum_{i=1}^{T} \gamma_k(x_i) \frac{(-1)^{1-x_i}}{\mu_{kd}(1-\mu_{kd})} \right) \frac{F_{\phi_{kd}}}{F_{\phi_{kd}}}
\]

where

\[
\gamma_k(x_i) = \frac{\pi_k p_k(x_i | \theta)}{\sum_{i=1}^{K} \pi_k p_k(x_i | \theta)}
\]

It is worth noting that the FV representation based on BMMs approach only takes into account the gradients with respect to \( \mu_{kd} \). Therefore, the \( KD \)-dimensional FV representation of a fingerprint sample is defined as \( \phi_B = [\phi_{kd}] \), \( k = 1, \ldots, K \) and \( d = 1, \ldots, D \).

Finally, the FV representation based on BMM yields a compact vector, whose size \( Kd \) is the half of FV encoding built upon GMM approach. In addition, BMM, unlike GMM, does not require data decorrelation (i.e. we do not need to apply PCA to the extracted local features).

3.2 Probabilistic visual vocabulary and Fisher vectors

As it was preliminarily explored in [21], 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 FV encoding derives a kernel from a generative model of the data (e.g. GMM [31], or BMM [32]). This in turn characterises how the distribution of a set of local descriptors, extracted from unknown or malicious PAIs, differs from the known PAI distribution previously learned by the adopted generative model. Therefore, the final transformed features are more robust to new samples, which may stem from unknown scenarios and thus differ from the samples used for training, as showed in a preliminary evaluation in [21].

In this article, we evaluate both continuous and binary texture descriptors (see Section 3.1 for more details). For continuous or real-valued descriptors (Figure 1, top), as proposed in [31], we train a GMM model with diagonal covariances from the fixed descriptors extracted on the previous step. In particular, a GMM on \( K \)-components, which is represented by their mixture weights (\( \pi_k \)), means (\( \mu_k \)) and covariance matrices (\( \sigma_k \)), 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 adopted descriptors are firstly decorrelated using principal component analysis (PCA) [52], 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 [30].

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

\[
FV_X = P(X | S_K) = P(X | \mu_k, \sigma_k)
\]

By applying Bayesian properties, we can rewrite previous equation as

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

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

where \( x_i(k) \) is the soft assignment weight of the \( i \)-th feature \( x_i \) to the \( k \)-th Gaussian. Finally, the FV representation that defines a fingerprint sample is obtained by stacking the differences \( \phi_G = [\phi_1^G, \phi_2^G, \ldots, \phi_K^G] \), thereby resulting a 2 \( \times \) \( 64 \cdot K \) size vector.

On the other hand, for encoding binary features (Figure 1, bottom) we train a BMM, whose \( K \)-components are represented by the mixture weights (\( \pi_k^B \)) and means (\( \mu_k^B \)), with \( k = 1 \ldots K \). Therefore, a closed-form approximation of FV representation is given by

\[
\phi_{\mu_{kd}} = \left( \sum_{i=1}^{T} \gamma_k(x_i) \frac{(-1)^{1-x_i}}{\mu_k(1-\mu_k)} \right) \frac{F_{\phi_{kd}}}{F_{\phi_{kd}}}
\]

\[
\gamma_k(x_i) = \frac{\pi_k p_k(x_i | \theta)}{\sum_{i=1}^{K} \pi_k p_k(x_i | \theta)}
\]

\[
F_{\phi_{kd}} = T \sum_{k=1}^{K} \pi_k \mu_{kd} + \sum_{k=1}^{K} \pi_k (1 - \mu_{kd}) (1 - \mu_{kd})^2
\]

Finally, the FV representation based on BMM yields a compact vector, whose size \( Kd \) is the half of FV encoding built upon GMM approach. In addition, BMM, unlike GMM, does not require data decorrelation (i.e. we do not need to apply PCA to the extracted local features).

3.3 Classification

Since that we perform a normalisation right after a signed square-rooting step over the FV representation, as indicated in [31], it can be efficiently used for learning a linear model. Hence, individual linear SVMs have been employed for each descriptor. SVMs are popular since they perform well in high-dimensional spaces, avoid over-fitting and have good generalisation...
capabilities. According to [53], when the feature’s dimensionality is so big in comparison with the number of instances used 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 for the bona fide samples as +1 and the attack presentations as −1, thereby yielding the corresponding \( W \) (weights) and \( b \) (bias) classifier parameters.

Subsequently, given a FV-based descriptor \( x \), the final score \( s_x \), which estimates the class of the sample at hand, is computed as the confidence of such decision (i.e. the absolute value of the score is the distance to the hyperplane):

\[
s_x = W \cdot x + b
\]

### 3.4 Fused approach

As mentioned above, and as it has been already shown in other pattern recognition, biometric, and PAD studies [10,12], the use of complimentary information can improve the recognition capabilities of a given approach. Therefore, we analyse to which extent different descriptors complement each other to improve the final PAD performance.

To that end, the individual descriptor-based PAD scores are fused with a weighted sum as follows:

\[
s_f = \alpha \cdot s_1 + \beta \cdot s_2 + (1 - \alpha - \beta) \cdot s_3,
\]

where \( \alpha + \beta < 1 \), and \( s_1, s_2 \) and \( s_3 \) represent the individual scores produced by the best three performing descriptors described above. Given that the LivDet databases do not include a validation set, the \( \alpha \) and \( \beta \) weighted values are computed from each LivDet’s training set.

### 4 EXPERIMENTAL SETUP

The experimental protocol has been developed in order to carry out a fair evaluation of the detection performance of the proposed PAD approach for different scenarios. Thus, the main aim of the evaluation is threefold: (i) analyse the impact of encoding different local descriptors on the detection performance, (ii) benchmark the performance of our proposal with the top state-of-the-art approaches, and (iii) evaluate realistic and challenging scenarios with unknown attacks, unknown capture device, and cross-database settings.

#### 4.1 Databases

The experiments were conducted on the well-known benchmarks provided by LivDet 2011 [54], LivDet 2013 [55], LivDet 2015 [56], and LivDet 2017 [57]. A summary of their main features is presented in Table 2. In addition, it should be noted that, unlike the previous databases, LivDet 2015 and LivDet 2017 contain unknown presentation attacks in the test set, which are not included in the training set. These unknown PAI species are highlighted in bold.

#### 4.2 PAD evaluation metrics

In order to perform an evaluation in compliance with the ISO/IEC 30,107-3 [33], we report the Attack Presentation Classification Error Rate (APCER), which denotes the percentage of misclassified presentation attacks for a fixed threshold, and the Bona Fide Presentation Classification Error Rate (BPCER), which refers to the percentage of misclassified bona fide presentations. Moreover, we include the corresponding Detection Error Trade-off (DET) curves between both detection errors, as well as the BPCER for a fixed APCER of 10% (BPCER10), 5% (BPCER20) and 1% (BPCER100). Finally, in order to allow a fair benchmark with the available literature, we report the ACER, which is computed as the average between the APCER and the BPCER for a fixed threshold \( \delta \in [0, 1] \).

#### 4.3 Experimental protocol

In all the experiments, we follow the LivDet evaluation protocol, where half of the database is used for training and another half for testing, and \( \delta = 0.5 \). In order to achieve the goals mentioned at the beginning of the Section 4, several scenarios are evaluated:

- **Baseline scenario**: in this baseline scenario, all testing and training samples for a specific evaluation are acquired with the same sensor, and all the materials used for building the PAIs are also known a priori in the training step. Therefore, this is the ideal optimal performance reached by any PAD method.

In this scenario, we first optimise the performance of the different texture descriptors over the training set. Subsequently, we select the best configuration for each particular descriptor and benchmark the corresponding detection performance with the state of the art.

We have followed the evaluation protocols proposed within the LivDet benchmarks. It is important to highlight that from the LivDet 2015 database, we only select in the test set PAI species known a priori from the training set. Furthermore, LivDet 2017 has only unknown attacks in the test set and hence is excluded from the baseline scenario.

- **Challenging scenarios**: in these scenarios, either the PAI species (i.e. unknown material scenario), or the sensors (i.e. unknown capture device and cross-database scenarios), are different from the train to the corresponding test sets. These approaches model more realistic scenarios, where the emergence of new spoof materials for
generating PAIs or new acquisition capture devices may lead to testing conditions in which the PAI species and/or capture devices are unknown at training time. Given the rapid evolution observed in the technology, information and communication industry, this is likely to occur in the next few years. Therefore, it is of utmost importance that our PAD methods are robust in these situations.

In contrast to the previous baseline scenario, the evaluations carried out under these scenarios provide a fair benchmark under challenging and realistic conditions of the PAD methods. In order to provide a better benchmark of the results, we follow the protocols presented in [18,56,57].

5 | EXPERIMENTAL RESULTS

5.1 | Baseline scenario: known materials and capture devices

5.1.1 | Effect of the semantic sub-groups

In the first step, in order to optimise the detection performance of the proposed method, different semantic sub-group sizes $K$ (i.e. the key parameter of the proposed method) are evaluated. In order to avoid bias due to other variables, we focused on the baseline scenario and tested different value ranges: $K = \{256, 512, 1024\}$, $K$ values greater than 1024 would result in larger feature vectors that are unusable for real-time applications and thus, are not considered in this work. Table 3 reports the ACER per descriptor for different $K$ values. As it should be observed, most descriptors report their best ACER for a small number of semantic sub-groups with the exception of BSIF and PHOG, which obtain their optimum performance for $K = 1024$. In particular, the SIFT descriptor yields an ACER of 2.23% for $K = 512$ which is up to three times lower than the ACER attained for $K = 1024$. This observation indicates that the FV representation is able to successfully represent the distribution of BPs and APs for a small set of semantic sub-groups. This also allows getting an efficiency improvement which can be exploited for real-time applications.

5.1.2 | Gradient versus texture versus intensity differences

It should be observed that the gradient-based descriptors report on average the best detection performance for all databases (i.e. ACER = 4.17%), followed by texture-based features (i.e. ACER = 5.12%), and finally, intensity difference-based descriptors (i.e. ACER = 7.42). By carefully analysing several PAI species from the LivDet databases, we noted that there exist at least five common artefacts which are fully represented by gradient- and texture-based descriptors and hence they could be employed for fingerprint PAD (see Figure 4). Specifically, the gradient computed over fingerprints allows representing their orientation field, hence capturing some ridge pattern characteristics such as, black and white saturation on the ridges, lack of continuity, ridge distortions or unwanted noises, non ridge uniformity, and spurious minutiae, among others which produce a high number of low coherence areas. Consequently, those ridge artefacts could be also captured by convolving a fingerprint image with a suitable kernel, as shown

---

**TABLE 2** Databases summary, including the list of PAI species available

| DB            | # Samples | Capture device          | PAI materials                                    |
|---------------|-----------|-------------------------|--------------------------------------------------|
| LivDet 2011   | 16,000    | Digital 4000B           | Gelatin, Latex, PlayDoh, Silicone, Wood Glue     |
|               |           | Sagem MSO300            |                                                  |
|               |           | Biometrika FX2000       | Gelatine, Latex, EcoFlex (platinum-catalysed silicone), Silgum, Wood Glue |
|               |           | Italdata ET10           |                                                  |
| LivDet 2013   | 8000      | Biometrika FX2000       | Gelatine, Latex, EcoFlex (platinum-catalysed silicone), Modasil, Wood Glue |
|               |           | Italdata ET10           |                                                  |
| LivDet 2015   | 19,431    | GreenBit DaetyScan26    | EcoFlex (platinum-catalysed silicone), Gelatin, Latex, Wood Glue, Liquid EcoFlex, RTV |
|               |           | Biometrika HiScan-PRO   |                                                  |
|               |           | Digital Persona U.are.U 5160 |                                                 |
|               |           | Crossmatch L Scan Guardian | PlayDoh, Body Double, EcoFlex, OOMOO (silicone rubber), novel form of gelatine |
| LivDet 2017   | 18,984    | Digital Persona U.are.U 5160 | EcoFlex, Body Double, Wood Glue, Gelatin, Latex, Liquid EcoFlex |
|               |           | GreenBit DaetyScan84 C  |                                                  |
|               |           | Orcanthus Certis2 image |                                                  |

Note: The unknown PAI species at testing for LivDet 2015 and LivDet 2017 are highlighted in bold.

Abbreviations: PAI, presentation attack instrument; DB, Database.
| DB          | Dataset   | SIFT | BSIF | SURF | PHOG | LBP | ORB | BRIEF |
|-------------|-----------|------|------|------|------|-----|-----|-------|
| 2011 Biometrika | 256 512 1024 | 2.80 | 4.10 | 5.70 | 2.40 | 2.80 | 2.60 | 4.00 5.30 4.90 15.40 | 15.50 | 12.80 | 7.80 | 6.90 | 7.70 | 7.50 | 6.20 | 6.20 | 11.60 | 11.00 | 10.70 | 11.70 |
| Digital P. | 0.70 | 0.30 | 0.30 | 0.60 | 0.50 | 0.40 | 1.50 | 1.20 | 1.10 | 2.30 | 1.00 | 0.30 | 1.10 | 1.10 | 1.60 | 4.20 | 3.40 | 3.70 | 1.30 | 1.30 | 1.20 |
| Itaklata | 3.20 | 2.40 | 4.50 | 6.50 | 6.90 | 6.10 | 16.60 | 13.10 | 13.40 | 19.70 | 20.30 | 20.00 | 15.60 | 19.30 | 20.90 | 8.30 | 9.90 | 11.60 | 16.10 | 15.60 | 15.60 |
| Sącem | 1.72 | 1.60 | 1.42 | 1.42 | 1.23 | 1.08 | 0.59 | 0.69 | 0.89 | 9.92 | 7.71 | 7.91 | 7.81 | 8.50 | 10.80 | 3.09 | 3.00 | 2.70 | 6.29 | 6.19 | 5.11 |
| 2013 Biometrika | 0.30 | 0.50 | 0.50 | 0.40 | 0.40 | 0.30 | 0.90 | 0.80 | 1.00 | 1.90 | 2.00 | 1.90 | 2.50 | 2.50 | 2.70 | 3.40 | 2.80 | 2.60 | 4.40 | 3.20 | 3.60 |
| Itaklata | 0.30 | 0.30 | 0.30 | 0.40 | 0.30 | 0.30 | 0.70 | 0.70 | 0.60 | 2.30 | 2.20 | 2.00 | 6.00 | 5.20 | 6.40 | 1.90 | 1.20 | 1.10 | 5.30 | 4.00 | 2.90 |
| 2015 GreenBit | 1.60 | 1.30 | 1.40 | 1.50 | 2.50 | 2.20 | 3.60 | 3.50 | 3.90 | 5.40 | 5.30 | 5.40 | 4.80 | 4.10 | 4.90 | 2.80 | 2.70 | 3.10 | 5.30 | 6.00 | 6.40 |
| Digital P. | 7.30 | 6.50 | 6.50 | 4.10 | 4.40 | 4.30 | 7.90 | 7.00 | 7.40 | 11.70 | 11.20 | 11.30 | 12.90 | 12.50 | 13.00 | 16.50 | 16.90 | 17.10 | 18.20 | 16.60 | 17.40 |
| Hi-Scan | 4.60 | 4.30 | 4.50 | 4.90 | 4.60 | 4.50 | 8.50 | 8.70 | 8.60 | 8.30 | 7.10 | 6.90 | 13.50 | 13.60 | 13.10 | 14.80 | 14.50 | 15.70 | 9.90 | 9.60 | 9.60 |
| Crossmatch | 1.06 | 1.03 | 1.03 | 2.47 | 2.12 | 2.22 | 8.34 | 7.53 | 7.62 | 2.59 | 1.88 | 2.09 | 6.34 | 6.21 | 6.59 | 9.40 | 9.40 | 10.46 | 5.41 | 5.06 | 5.03 |
| Avg. | 2.36 | 2.23 | 2.62 | 2.47 | 2.58 | 2.40 | 5.24 | 4.85 | 4.94 | 7.95 | 7.44 | 7.06 | 7.84 | 7.99 | 8.77 | 7.19 | 7.00 | 7.43 | 8.38 | 7.83 | 7.85 |

Note: Detection performance, in terms of ACER (%), of our proposed descriptors for different $K$ values. The best results per descriptor are highlighted in bold, and the best ACER per dataset is underlined.

Abbreviations: BRIEF, Binary Robust Independent Elementary Features; BSIF, Binarised Statistic Invariant Features; HOG, Histogram of Oriented Gradients; LBP, Local Binary Pattern; ORB, Oriented FAST and Rotated BRIEF; PHOG, Pyramid HOG; SIFT, Scale Invariant Feature Transform; SURF, Speed-Up Robust Features; DB, Database.
in Figure 4 third row. As it was mentioned, in this work 60 filter kernels are employed for the BSIF computation. The best performing filter configurations per LivDet dataset are reported in Table 4. As it should be noted, a texture-based descriptor as BSIF achieves its best detection performance for small-size filter kernels at most cases (i.e. \(N \leq 9\)); large-size filter kernels can lead to a deterioration of the fingerprint ridge pattern structure, thereby removing the aforementioned artefacts. Finally, we can observe in Table 3 that intensity difference-based features analysed by ORB and BRIEF are not suitable to detect an attack presentation attempt, thereby resulting in a poor detection performance.

### 5.1.3 Effect of fingerprint quality

Keeping in mind that fingerprint ridge patterns between a BP and AP could be different, we computed the NFIQ2.0 [58] quality of samples in LivDet 2015 for the best descriptor category (i.e. Gradient, Texture and Intensity difference). Figure 5 reports the detection performance behaviour of the adopted descriptor categories as the fingerprint ridge pattern of bona fide samples improves. As it can be noted, all descriptor categories achieve a detection performance improvement as the bona fide ridge pattern quality increases. In particular, gradient-based descriptors yield a mean ACER of 2.36% for bona fide images with a NFIQ2.0 quality greater than 40, which outperforms the texture- and intensity difference-based features by a relative 20% and 73%, respectively. These findings in turn confirm the soundness of the gradient-based descriptors to capture the aforementioned ridge pattern artefacts, hence detecting the attack presentation attempts.

**TABLE 4** Best performing BSIF filter configurations per dataset and \(K\) value. The best results per dataset are highlighted in bold

| DB       | Dataset | BSIF Parameters |
|----------|---------|-----------------|
| Biometrika | N = 5, l = 5 | N = 5, l = 5 | N = 5, l = 5 |
| Digital P. | N = 11, l = 7 | N = 11, l = 7 | N = 11, l = 7 |
| Italdata | N = 7, l = 11 | N = 11, l = 5 | N = 8, l = 3 |
| Sagem | N = 8, l = 7 | N = 8, l = 7 | N = 8, l = 7 |
| Biometrika | N = 6, l = 3 | N = 6, l = 3 | N = 6, l = 5 |
| Italdata | N = 10, l = 5 | N = 6, l = 9 | N = 9, l = 7 |
| 2015 GreenBit | N = 7, l = 5 | N = 7, l = 3 | N = 7, l = 3 |
| Digital P. | N = 6, l = 3 | N = 6, l = 3 | N = 6, l = 3 |
| Hi-Scan | N = 7, l = 3 | N = 7, l = 3 | N = 7, l = 3 |
| Crossmatch | N = 8, l = 9 | N = 12, l = 9 | N = 12, l = 9 |
| 2017 Digital P. | N = 12, l = 7 | N = 12, l = 7 | N = 12, l = 7 |
| GreenBit | N = 7, l = 3 | N = 7, l = 3 | N = 7, l = 3 |
| Oreamthus | N = 6, l = 3 | N = 6, l = 3 | N = 7, l = 5 |

Abbreviation: BSIF, Binarised Statistic Invariant Features; DB, Database.
5.1.4 | State-of-the-art benchmark

Finally, Table 5 establishes a benchmark, in terms of ACER, of the best performing three descriptors (i.e. SIFT, BSIF, and SURF) against the current top state of the art. In addition, a score-level fusion between them is computed. In our work, we also experimented with the fusion between the best descriptors per category (i.e. SIFT, BSIF, and ORB). However, the low discriminative power of the intensity differences-based descriptor led to a clear performance deterioration for unknown attacks. As it may be observed, most analysed descriptors as well as their fusion achieve the state of the art for most datasets. In particular, the proposed fusion reports, on average, remarkable ACERs of 0.95%, 0.30%, and 1.46% for the three LivDet databases. In addition, we can see that, in contrast to most Deep learning approaches, our fusion method is able to yield a good detection performance for Digital Persona in LivDet 2015 (i.e. ACER of 0.10%). According to Ghiani et al. [59], most algorithms submitted to LivDet 2015 did not perform well on Digital Persona due to the small image size. Moreover, Gonzalez-Soler et al. [22] showed some NFIQ2.0 quality findings which indicate that some capture devices, namely Digital Persona U.are.U5160 and Biometrika Hi-Scan-PRO, include an acquisition technology that produces a high unwanted noise degree on the fingerprint ridge pattern. This fact affects in turn the detection performance of most state-of-the-art methods [18,42,43].

5.1.5 | In-depth detection performance analysis

In order to analyse the usability of our fusion method for an operational real application, we evaluate its detection performance in compliance with the ISO/IEC 30,107-3 [33] in Figure 6. As it may be observed, the performance varies considerably from the best (i.e. Digital Persona subset from LivDet 2015, with a BPCER = 0% for APCER ≥ 0.5% or APCER = 0% for BPCER ≥ 0.05%) to the worst case (i.e. Hi-Scan subset from LivDet 2015, with a BPCER100 = 10.60%). Specifically, our fusion-based algorithm reports on average a BPCER100 of 1.48% and 0.05% for LivDet 2011 and LivDet 2013, respectively, which are up to seven times lower than the ones attained by the current state-of-the-art techniques (i.e. a BPCER100 of 9.68% for FSB-v1 [17] and 4.05% for FSB-v2 [16] on the LivDet 2011, and a BPCER100 of 0.20% for FSB-v1 [17] and 0.05% for FSB-v2 [16] on the LivDet 2013). Consequently, the proposed fusion is able to achieve a top detection performance for most datasets in LivDet 2015: a mean BPCER100 of 0.63% for the Crossmatch, GreenBit, and

![FIGURE 5 Average classification error rate benchmark in terms of NFIQ2.0 quality per descriptor category](image)

**TABLE 5** Benchmark in terms of the ACER (%) against the top state-of-the-art

| DB       | Dataset          | VGG [18] | TripleNet [20] | FSB-v1 [17] | tinyFCNN [42] | FSB-v2 [16] | FLDNet [43] | LocalEnc [22] | SIFT | BSIF | SURF | FUSION |
|----------|------------------|---------|----------------|-------------|---------------|-------------|-------------|---------------|------|------|------|--------|
| 2011     | Biometrika       | 5.20    | 5.15           | 2.60        | **1.10**      | 1.24        | -           | 2.40          | 2.80 | 2.40 | 4.00 | **1.50** |
|          | Digital P.       | 3.20    | 1.85           | 2.70        | 1.10          | 1.61        | -           | **0.10**      | 0.30 | 0.40 | 1.10 | **0.10** |
|          | Italdata         | 8.00    | 5.10           | 3.25        | 4.75          | 2.45        | -           | 2.20          | 2.40 | 6.10 | 13.10 | **1.90** |
|          | Sagem            | 1.70    | 1.23           | 1.80        | 1.56          | 1.39        | -           | 1.13          | 1.42 | 1.08 | 0.59 | **0.29** |
|          | **Avg.**         | **4.52**| **3.33**       | **2.59**    | **3.12**      | **1.67**    | -           | **1.46**      | 1.73 | 2.50 | 4.70 | **0.95** |
| 2013     | Biometrika       | 1.80    | 0.65           | 0.60        | 0.35          | **0.20**    | **0.36**    | 0.30          | 0.30 | 0.30 | 0.80 | 0.30 **(α = 0.0, β = 0.6)** |
|          | Italdata         | 0.40    | 0.50           | 0.40        | 0.40          | **0.30**    | 1.35        | **0.30**      | **0.30** | **0.30** | **0.60** | **0.30** **(α = 0.0, β = 0.8)** |
|          | **Avg.**         | **1.10**| **0.58**       | **0.50**    | **0.38**      | **0.25**    | **0.86**    | **0.30**      | **0.30** | **0.30** | **0.70** | **0.30** |
| 2015     | GreenBit         | 4.60    | -              | 2.00        | **0.20**      | 0.68        | 0.53        | 1.30          | 1.30 | 1.50 | 3.50 | 1.00 **(α = 0.3, β = 0.5)** |
|          | Digital P.       | 5.64    | -              | 1.76        | 3.40          | 1.12        | 3.61        | 6.20          | 6.50 | 4.10 | 7.00 | **0.10** **(α = 0.6, β = 0.1)** |
|          | Hi-Scan          | 6.28    | -              | 1.08        | **0.35**      | 1.48        | 2.95        | 4.30          | 4.30 | 4.50 | 8.50 | 3.80 **(α = 0.5, β = 0.3)** |
|          | Crossmatch       | 1.90    | 0.81           | 1.09        | **0.64**      | 1.78        | 1.03        | 1.03          | 2.12 | 7.53 | 0.94 **(α = 0.1, β = 0.8)** |
|          | **Avg.**         | **4.61**| **1.39**       | **1.26**    | **0.97**      | **2.22**    | **3.20**    | **3.28**      | **3.06** | **6.63** | **1.46** |

Note. Benchmark in terms of the ACER (%) against the top state-of-the-art. The best result per dataset is highlighted in bold.

Abbreviations: ACER, average classification error rates; BSIF, Binarised Statistic Invariant Features; SIFT, Scale Invariant Feature Transform; DB, Database.
Digital Persona datasets (i.e. state-of-the-art: 1.61% in FSB-v1 [17] and 0.92% in FSB-v2 [16]). In addition, it should be noted that our method yields its worst detection performance for the Hi-Scan capture device, thereby resulting in a BPCER100 of 10.60%. The Hi-Scan dataset includes high-resolution fingerprints with sizes of 1000 × 1000 pixels where the ROI for most PAIs only covers a 40% of a whole image. In contrast, the ROI for BP samples covers up to 70% of pixels in the images. Since our proposed approach extracts the descriptors from the whole image, we think that a ROI segmentation or a reduction of the points on the regular grid to particular landmarks such minutiae, for the feature extraction, could lead to a detection performance improvement for this type of high-resolution capture device.

Finally, it should be noted that, in general, a good balance between high user convenience or usability (i.e. low BPCER) and high security (i.e. low APCER) can be achieved with the proposed method. In particular, the BPCER ranges between 0.12% and 1.85% for higher security thresholds (i.e. 1.0% ≤ APCER ≤ 10.0%) confirming the remarkable detection performance of the fusion between gradient and texture-based descriptors for this baseline scenario.

5.2 | Challenging scenarios: unknown PAI species and/or capture devices

As mentioned in Sect. 4.3, the following experiments evaluate the generalisation capabilities of the proposed PAD schemes to identify unknown attacks for different scenarios. In order to show the usability of our fusion approach for a realistic scenario, we considered the sub-optimal fusion weights, the best performing K values, and fixed descriptor's parameters obtained in Sect. 5.1.

5.2.1 | Unknown material scenario

In order to evaluate the generalisation capability of our three best performing descriptors and their fusion to detect PAIs fabricated with unknown artificial materials, two experiments are performed following three different protocols. In both experiments, all training and test images were acquired by the same capture device.

In the first set of experiments, we select the LivDet 2015 database, in which the unknown PAI species employed for the fabrication of the Crossmatch test AP images include Gelatin and OOMOO, while PAI species such as RTV and Liquid EcoFlex are included in the Digital Persona, GreenBit and Hi-Scan test sets. For the unknown material scenario within the LivDet 2011 and 2013 databases, we follow the experimental protocol described in [18]. Following this idea, an unknown material evaluation is also carried out over the LivDet 2017. Table 6 shows the corresponding ACER values for all subsets.

Focusing first on the LivDet 2011 and LivDet 2013 databases for the unknown material scenario, a similar trend to the baseline scenario can be observed for the three selected local descriptors: the gradient-based descriptors achieve on average the best performance for most datasets (average ACER = 2.08% for SURF), followed by the texture-based descriptor (ACER = 2.36% for BSIF). In addition, it should be observed that the three descriptors and their fusion outperform the top state-of-the-art. In particular, the fusion method yields a mean ACER of 1.00%, which is approximately 3 and 10 times better than the ones attained by the best methods. These results can be also be observed for the LivDet 2017 database, reporting on average an ACER of 3.97% which is better than the one attained by the LivDet 2017 winner [57].

Regarding the experiments run on LivDet 2015 (see Table 6, mid row), the trend observed for LivDet 2011 and LivDet 2013 is confirmed: gradient-based descriptors show the best performance, followed by the texture-based one (i.e. 5.41% vs. 5.61%). On the other hand, it may be noted that our fusion method suffers a high performance deterioration for Digital Persona and Hi-Scan due to their fingerprint quality; 90% and 70% of fingerprint images in those datasets report a NFIQ 2.0 quality below 50%, thereby producing an accuracy decrease for most PAD techniques [18,42,43]. Therefore, those images are unsuitable to the PAD task and hence for a real fingerprint recognition system.

To conclude the analysis of this scenario, the ISO compliant evaluation of the fusion approach over all datasets is presented in Figure 7. As it was expected, the performance is worse than that over the baseline scenario (see Figure 6). Nevertheless, an average BPCER100 of 8.30%, BPCER20 of 3.47%, and BPCER10 of 1.93% can be achieved, thus still granting a secure and usable system. In addition, it should be noted that the gap in performance at the ACER for Digital Persona and Hi-Scan in the LivDet 2015 database is here...
| Protocol | Dataset | Train Species | Test Species | SIFT | BSIF | SURF | FUSION | LocalEnc [22] | FSB-v2 [16] | FLDEnt [43] | LivDet 2017 Winner [57]* |
|----------|---------|----------------|--------------|------|------|------|--------|---------------|-------------|---------------|-----------------------------|
| Proposed by Bio11 | EcoFlex, Gelatine, Latex | Silgum, Woodglue | 6.33 | 3.72 | 4.61 | **1.73** | 4.78 | 4.60 | - | - |
| [18] Bio13 | Modasil, Woodglue | EcoFlex, Gelatine, Latex | 1.00 | **0.87** | 1.31 | 1.04 | 1.50 | 1.30 | **0.87** | - |
| Ita11 | EcoFlex, Gelatine, Latex | Silgum, Woodglue, other | 3.78 | 4.61 | 2.00 | **1.11** | 3.60 | 5.20 | - | - |
| Ita13 | Modasil, Woodglue | EcoFlex, Gelatine, Latex | 0.30 | 0.23 | 0.41 | **0.13** | 0.50 | 0.60 | 0.94 | - |
| Average | | | 2.85 | 2.36 | 2.08 | **1.00** | 2.60 | 2.93 | - | - |
| LivDet 2015 | Crossmatch | Body Double, EcoFlex, PlayDoh | Gelatine, OOMOO | 1.37 | 2.56 | 6.87 | 2.01 | **1.34** | - | 2.66 | - |
| Digital P. | EcoFlex, 00-50, Latex, Gelatine, Woodglue | Liquid EcoFlex, RTV | 9.40 | 5.80 | 10.00 | 8.45 | 8.85 | - | **3.06** | - |
| GreenBit | | | 4.20 | 5.45 | 7.80 | 4.65 | 4.20 | - | **0.46** | - |
| HiScan | | | 6.65 | 8.65 | 11.60 | 6.00 | 6.65 | - | **3.38** | - |
| Average | | | 5.41 | 5.61 | 5.97 | 5.28 | 5.26 | - | **2.39** | - |
| LivDet 2017** | Digital P. | Woodglue, EcoFlex, Body Double | Gelatine, Latex, Liquid EcoFlex | 5.11 | 5.97 | 5.03 | **4.22** (α = 0.3, β = 0.4) | 4.84 | - | - | 4.41 |
| GreenBit | | | 5.64 | 5.35 | 5.88 | 4.06 (α = 0.3, β = 0.4) | 5.35 | - | - | **3.56** |
| Orcanthus | | | 6.16 | 4.04 | 7.00 | **3.63** (α = 0.6, β = 0.2) | 5.62 | - | - | 6.29 |
| Average | | | 5.64 | 5.12 | 5.97 | **3.97** | 5.15 | - | - | 4.75 |

Note: Detection performance of our best descriptors, in terms of ACER (%), for several unknown material scenarios. The best result per dataset is highlighted in bold.

*The overall classification errors reported by LivDet 2017 winner in this work are the complement of the overall accuracy achieved in [57].
**The ACER results were achieved at K = 512 for SIFT, and K = 256 for BSIF and SURF.
confirmed for all operating points (i.e. a BCPER100 of 19.60% for Hi-Scan and a BPCER100 of 29.30% for Digital Persona).

5.2.2 | Unknown capture device scenario

We now evaluate the unknown capture device scenario, where different capture devices might be used for training and testing at some point in time. This scenario is likely to happen in a long-time deployment, where the fingerprint capture device might age and eventually stop working. Therefore, the fabrication and acquisition of the entire set of earlier known PAI species with the new capture device at hand might not be possible or at least require some time, thereby being not available for high-security applications.

In order to study the generalisability of our proposed method for unknown capture devices, we adopt four training set—test set configurations proposed by [18]. Table 7a reports the corresponding ACER values. As it may be observed, a gradient-based descriptor (i.e. SIFT) still provides the lowest error rates (ACER = 4.08%). However, it should be noted that BSIF deploys a similar detection performance for the same peerwise configurations. Even if Italdata 2011 and Biometrika 2011 visually show different texture patterns (see Figure 8), BSIF is able to yield similar ACER values when these datasets are used for testing (i.e. ACER of 11.65% for Bio11-Ital11 vs. 10.05% for Ital11-Bio11). Consequently, a similar behaviour can be observed for Biometrika 2013 and Italdata 2013. However, they visually look more similar, thereby resulting in a low ACER for all descriptors. Therefore, in order to successfully achieve the interoperability requirement between capture devices, the selection of a new sensor must be carefully performed taking into account the five fingerprint ridge pattern properties mentioned in Sect. 5.1.2. This could in turn avoid a

![Figure 7](image)

**FIGURE 7** Detection error trade-off curves on the unknown or unknown material scenario for the proposed fusion method

**TABLE 7** ACER evaluation for challenging scenarios

|          | SIFT | BSIF | SURF | FUSION | LocalEnc [22] | FSB-v2 [16] | FLDNet [43] |
|----------|------|------|------|--------|---------------|-------------|-------------|
| Bio11 - Ital11 | 11.30 | 11.65 | 11.40 | 9.65 | 11.45 | 25.35 | - |
| Bio13 - Ital13 | 1.80 | 2.95 | 0.95 | 1.55 | 1.80 | 4.30 | 2.10 |
| Ital11 - Bio11 | 2.40 | 10.05 | 7.95 | 3.70 | 7.40 | 25.21 | - |
| Ital13 - Bio13 | 0.80 | 3.30 | 1.35 | 0.90 | 0.75 | 3.50 | 2.90 |
| Avg | 4.08 | 6.99 | 5.41 | 3.95 | 5.35 | 14.59 | - |

|          | SIFT | BSIF | SURF | FUSION | LocalEnc [22] | FSB-v2 [16] | FLDNet [43] |
|----------|------|------|------|--------|---------------|-------------|-------------|
| Bio11 - Ital13 | 6.80 | 3.95 | 3.90 | 5.90 | 4.00 | 7.60 | 14.00 |
| Bio13 - Ital11 | 12.70 | 16.95 | 18.55 | 14.75 | 13.60 | 31.16 | 34.05 |
| Ital11 - Ital13 | 5.60 | 2.95 | 9.40 | 6.95 | 5.60 | 6.70 | 8.30 |
| Ital13 - Ital11 | 11.50 | 21.95 | 20.10 | 24.10 | 17.50 | 26.16 | 44.65 |
| Avg | 9.15 | 11.45 | 12.99 | 12.92 | 10.18 | 17.91 | 25.25 |

Note. ACER (\%) evaluation for challenging scenarios. The best result per dataset is highlighted in bold.
gap on the detection performance of the state-of-the-art techniques.

Regarding the fused scheme, for this scenario our approach considerably outperforms the state of the art by a relative 73% (i.e. ACER = 3.95% vs. 14.59%), thereby showing its generalisation capability for this challenging scenario.

In Figure 9-left, the ISO compliant evaluation of the fusion scheme is depicted. We can first observe the increased detection performance gap between LivDet 2011 and LivDet 2013 for all operating points. In particular, our approach yields a mean BPCER 100 of 2.40% for datasets in LivDet 2013, which is approximately nine times lower than the one reported for datasets in LivDet 2011 (i.e. BPCER = 18.18%). Despite this detection performance gap, our fused method is able to achieve a mean BPCER100 of 10.14%, thereby providing both user convenience and security.

5.2.3 | Cross-database scenario

Finally, we evaluate the scenario where different data collection sessions for the same capture device are used for training and testing. To that end, we select two datasets (i.e. Biometrica and Italdata), whose sensors were respectively used for fingerprint acquisition in the LivDet 2011 and LivDet 2013 competitions. Table 7b shows the corresponding ACER values.

As in most scenarios analysed, a gradient-based descriptor (i.e. SIFT) provides the best detection performance. In particular, SIFT reports a mean ACER below 10%, which outperforms the remaining descriptors, their fusion and the top state-of-the-art techniques. It should be noted that our descriptors and their fusion suffer a detection performance deterioration when the LivDet 2011 dataset is employed for testing. However, each dataset captured at different years are visually similar (see Figure 8), they report a different detection performance. Specifically, the evaluation of LivDet 2011 attains a mean ACER of 16.95%, in contrast to 5.43% reported by testing LivDet 2013. As for unknown material results, the inclusion of an unknown material as Silgum in LivDet 2011 is one of the issues leading to a decrease in accuracy of our approach. Finally, Figure 9 (right) confirms the detection performance gap over the LivDet 2011 and 2013 databases (i.e. blue and grey vs. yellow and red): a higher BPCER100 over 60% for LivDet 2011 yields a non-userable fingerprint systems.

5.3 | Computational efficiency

In the last experiments, we evaluate the computation efficiency of our fusion-based method for $K = 1024$ (i.e. the worst case in terms of computational load) on an Intel Core i7-8750H @ 2.2 GHz, 16 GB RAM. To that end, we select the LivDet 2015 datasets having the largest images and compute the average classification time for each particular descriptor (i.e. SIFT, BSIF, and SURF). As a result of the experiments, the best three performing descriptors report an average time of 1.51, 1.47, 1.12 s, respectively, to analyse a probe sample. Since the three pipelines can be simultaneously executed, the computational efficiency of our fusion-based approach is bounded by the time yielded by the most time-consuming descriptor (i.e. 1.51 s for SIFT). Given that the computational cost of our algorithm increases with the number of points on the regular grid employed for the feature extraction, their reduction to landmarks such as the minutiae would significantly benefit its efficiency without losing accuracy. This efficiency limitation will be addressed in the future work.
5.4 | Visualisation of the FV representation

Finally, we show in Figure 10 the visualisation of the scores predicted by our approach for misclassified and correctly classified samples taken from the LivDet 2015. It should be noted that both the heatmaps for misclassified BPs and correctly classified APs contain a high number of low coherence areas or unwanted noise, in contrast to the ones yielded for wrongly classified APs and correctly classified BPs. Those areas of low coherence are produced by the capture devices in the sample acquisition. In addition, we may observe that our proposed method fails for those PAs having a perfectly defined ridge pattern, as depicted in Figure 10a. As it was mentioned above, local analysis for particular landmarks such as minutiae could lead to an improvement for these challenging cases.

5.5 | Summary

To summarise the findings of the experimental evaluation, we can highlight the following take-away messages:

- The proposed PAD method, based on local image descriptors and feature encoding, is able to outperform the state of the art not only in the baseline scenario (i.e. both the PAs and the acquisition devices are known a priori) but also in more realistic and challenging scenarios (i.e. unknown material, unknown capture device and cross-database). In particular, the ACER is reduced by up to four times for the unknown capture device scenario (i.e. ACER = 3.95% vs. 14.59%).
- In addition, the ISO compliant evaluation revealed that the fused approach provides an usable system (i.e. low BPCER) even for a high security (APCER = 1%) operating point; it achieves an average BPCER100 = 1.85% for the baseline scenario; BPCER100 < 1.20% for LivDet 2013, BPCER100 < 16% LivDet 2015 and BPCER100 < 9% LivDet 2017 for the unknown material scenario; and BPCER100 < 11% for the unknown capture device evaluation.
- Gradient-based descriptors (i.e. SIFT and SURF) successfully represent low coherence areas produced by several fingerprint ridge pattern artefacts such as black saturation, white saturation, lack of continuity, unwanted noises and ridge distortions, thereby resulting in the best detection performance in most scenarios.
- Texture-based features also yield the best detection performance right after gradient-based descriptors. In particular, BSIF achieves its best performance for small filter sizes (i.e. N < 9), which capture most of aforementioned artefacts. Given that BSIF depends on a set of filters previously learnt from 13 natural images, we think that the use of filters trained for the particular fingerprint PAD task or extracted from intermediate CNN layers could unveil other ridge artefacts, hence improving the BSIF performance.
- Even if the Intensity differences- or binary-based descriptors (i.e. ORB, BRIEF) offer a lower computational load, their performance is not competitive against their continuous counterparts for fingerprint PAD purposes (i.e. SIFT, SURF, BSIF, PHOG and LBP).
- The fusion of gradient- and texture-based information considerably improves the detection performance of the single descriptors, even in scenarios where textural features alone achieve considerably higher error rates (e.g. unknown capture device).
- The semantic sub-groups learned by the GMM allow modelling the most aforementioned artefacts produced in the creation of PAs. A better artefact description by the semantic sub-groups depends on the input features that follow a Gaussian distribution. In order to remove this GMM constrain and hence improve the FV representation, new deep generative models, which have shown to be more powerful for learning data distribution, could be evaluated.
- Although deep-learning-based fingerprint PAD approaches require large databases for optimising thousands of parameters, our proposal attained a high detection performance training a small number of them (K, α, and β) from a small dataset.
- Furthermore, most state-of-the-art techniques yield a poor detection performance over the Digital Persona in LivDet
2015 due to its small image size, and our fusion method reports a remarkable ACER of 0.10%.

- Since that Hi-Scan contains images with a high size of 1000 \times 1000 pixels where the ROI area for PAIs only covers a 40% of the whole image, our fusion-based representation is unable to report a reliable detection performance, thereby resulting in a BPCER100 of 10.60%. A reduction of the points on the regular grid to specific landmarks such as minutiae, for the feature extraction, or a ROI segmentation could improve its error rates.

- A NFIQ2.0 evaluation over the LivDet 2015 database showed that different analysed descriptors improved their detection performance so the ridge pattern of bona fide fingerprints enhanced. Therefore, the NFIQ2.0 can be employed as a secure indicator in order to obtain a reliable PAD module; an ACER <2.58% is reported when the BP fingerprint quality is greater than 60 (i.e. NFIQ2.0 > 60).

### 6 CONCLUSIONS

In this article, we proposed a new PAD approach based on the FV encoding of different common image local descriptors, namely: SIFT, SURF, BSIF, LBP, PHOG, ORB and BRIEF. Moreover, a fusion of both gradient (SIFT, SURF) and texture (BSIF) descriptors have been analysed in order to improve the classification capability of the proposed algorithm. The experimental evaluation conducted over the publicly available LivDet 2011, LivDet 2013, LivDet 2015, and LivDet 2017 benchmarks assessed the performance of our proposals against the top state-of-the-art methods.

The extensive experimental evaluation revealed some differences across local descriptors for PAD purposes: gradient-based features are more robust than the remaining descriptors. Therefore, it successfully capture low coherence areas produced in the fabrication of PAIs. On the other hand, the fusion of gradient- and texture descriptors increased the detection performance in most scenarios, as it could be expected.

In more details, the proposed fused approach outperformed the top state-of-the-art results [16,17,20,43] in all scenarios. In particular, the relative improvement of the ACER reached 73% for one of the most challenging scenarios—using different capture devices for training and testing. On the other hand, our FV representation reports its worst results for high-resolution fingerprint images. A reduction of points on the regular grid to landmarks such as minutiae, for the feature extraction, could produce an improvement in its detection performance for this type of fingerprint images.

The aforementioned results joined with the fact that the SIFT-based configuration reported the best detection performance in the LivDet 2019 competition, with an overall accuracy of 96.17%, show the high generalisation capability of the handcrafted feature representation based on Fisher Vector in the task of PAD.

Given that our final representation depends on a generative model which learns the distribution of local descriptors, we plan to evaluate other generative models which have shown to be more powerful than the GMM to learn data distribution. We would thus improve the generalisation capability of our FV-based approach and hence remove the GMM assumption on the input data distribution—data following a Gaussian distribution lead to the building of better semantic sub-groups.

### ACKNOWLEDGEMENTS

This research work has been funded by the DFG-ANR RESPECT Project (406880674), and the German Federal Ministry of Education and Research and the Hessian Ministry of Higher Education, Research, Science and the Arts within their joint support of the National Research Center for Applied Cybersecurity ATHENE.

### ORCID

Lázaro J. González-Soler [https://orcid.org/0000-0001-6470-2966]

Marta Gomez-Barrero [https://orcid.org/0000-0003-4581-5533]

Jascha Kolberg [https://orcid.org/0000-0002-3128-8049]

Leonardo Chang [https://orcid.org/0000-0002-0703-2131]

Airel Pérez-Suárez [https://orcid.org/0000-0001-8796-6243]

Christoph Busch [https://orcid.org/0000-0002-9159-2923]

### REFERENCES

1. ISO/IEC JTC1 SC37 Biometrics: Information Technology - Biometric Presentation Attack Detection - Part 1: Framework. International Organization for Standardization (2016). ISO/IEC 30107-1

2. Soudénik, C., Buseh, C.: Presentation attack detection methods for fingerprint recognition systems: a survey. IET Biometr. 3(1), 1–15 (January 2014)

3. Marasco, E., Ross, A.: A survey on anti-spoofing schemes for fingerprint recognition systems. ACM Comput. Surv. 47(2), 28 (2015)

4. Galbally, J., Gomez-Barrero, M.: Presentation attack detection in iris recognition. In: Iris and Periocular Biometrics (2017)

5. Galbally, J., Marcel, S., Fierrez, J.: Biometric anti-spoofing methods: a survey in face recognition. IEEE Access. 2, 1530–1552 (December 2014)

6. Drahansky, M.: Experiments with skin resistance and temperature for liveness detection. In: Proc. International Conference on Intelligent Information Hiding and Multimedia signal Processing, pp. 1075–1079 (2008)

7. Reddy, P.V. et al.: A new anti-spoofing approach for biometric devices. IEEE Trans. Biomed. Circuits Syst. 2(4), 328–337 (2008)

8. Lapsley, P.D. et al.: Anti-fraud biometric scanner that accurately detects blood flow. US Patent. 5, 439–737 (1998)

9. Kolberg, J. et al.: Presentation attack detection for finger recognition. In: Handbook of Vascular Biometrics, pp. 435–463. Springer (2020)

10. Keilbach, P. et al.: Fingerprint presentation attack detection using laser speckle contrast imaging. Proc. International Conference on the Biometrics Special Interest Group (BIOSIG) (September 2018)

11. Baldissera, D. et al.: Fake fingerprint detection by odor analysis. In: Advances in Biometrics: International Conference (ICB 2006), pp. 265–272. Springer Berlin/Heidelberg (2006)

12. Tolosana, R. et al.: Biometric presentation attack detection: beyond the visible spectrum. IEEE Trans. Inf. Forensics Secur. 15, 1261–1275 (2019)

13. Gomez-Barrero, M., Kolberg, J., Busch, C.: Multi-modal fingerprint presentation attack detection: Looking at the surface and the inside. Proc. International Conference on Biometrics (ICB) (2019)

14. Nikam, S.B., Agarwal, S.: Ridgelet-based fake fingerprint detection. Neurocomputing. 72(10), 2491–2506 (2009)

15. Yambay, D. et al.: Review of fingerprint presentation attack detection competitions. In: Handbook of Biometric Anti-Spoofing, pp. 109–131. Springer (2019)
16. Chugh, T., Cao, K., Jain, A.K.: Fingerprint spoof buster: use of minutiae-centered patches. IEEE Trans. Inforn. Forensics Secur. 13(9), 2190–2202 (2018)

17. ———: Fingerprint spoof detection using minutiae-based local patches. In: Proc. International Joint Conference on Biometrics (IJCB), pp. 581–589 (2017)

18. Nogueira, R.F., de Alencar Lotufo, R., Machado, R.C.: Fingerprint liveness detection using convolutional neural networks. IEEE Trans. Inforn. Forensics Secur. 11(6), 1206–1215 (2016)

19. Chugh, T., Jain, A.K.: Fingerprint spoof detector generalization. IEEE Trans. Inforn. Forensics Secur. (2020)

20. Pala, F., Bhanu, B.: Deep triplet embedding representations for liveliness detection. In: Deep Learning for Biometrics, pp. 287–307 (2017)

21. González-Soler, L.J., et al.: Fingerprint presentation attack detection method based on a bag-of-words approach. In: Proc. Iberoamerican Congress on Pattern Recognition, (CIARP), pp. 263–271 (2017)

22. González-Soler, L.J., et al.: Fingerprint presentation attack detection based on local features encoding for unknown attacks. IEEE Access. 9, 5806–5820 (January 2021)

23. Lowe, G.D.: Distinctive image features from scale-invariant keypoints. Int. J. Comput. Vis. 60, 91–110 (2004)

24. Kannala, J., Rahlu E.: BSIF: Binarized statistical image features. In 21st International Conference on Pattern Recognition (ICPR), pp. 1363–1366 (2012)

25. Ojala, T., Pietikäinen, M., Maenpää, T.: Multiresolution gray-scale and rotation invariant texture classification with local binary patterns. IEEE Pattern Anal. Mach. Intell. 24(7), 971–987 (July 2002)

26. Dalal, N., Triggs, B.: Histograms of oriented gradients for human detection. In: 2005 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR'05), pp. 886–893.IEEE (2005)

27. Bay, H., Tuytelaars, T., Gool, L.V.: Surf: Speeded up robust features. In: Proc. International European Conference on Computer Vision (ECCV), pp. 404–417 (2006)

28. Calonder, M. et al.: Brief: binary robust independent elementary features. In Proc. International European Conference on Computer Vision, pp. 778–792.Springer (2010)

29. Rublee, E. et al.: An efficient alternative to sift or surf. Proc. Intl. Conf. Comput. Vis, 2564–2571 (2011)

30. Simonyan, K. et al.: Fisher vector faces in the wild. BMVC. 2(3), 4 (2013)

31. Sánchez, J. et al.: Image classification with the Fisher vector: Theory and practice. Int. J. Comput. Vis. 105(3), 222–245 (2015)

32. Uchida, Y., Sakazawa, S., Satoh, S.: Image retrieval with Fisher vectors of binary features. IET Trans. Media Technol. Appl. 4(4), 326–336 (2016)

33. ISO/IEC JTC1 SC37 Biometrics: Information Technology – Biometric Presentation Attack Detection—Part 3: Testing and Reporting. International Organization for Standardization (2017). ISO/IEC 30107-3

34. Xia, Z. et al.: Combining wavelet transform and lbp related features for fingerprint liveliness detection. IAENG Int. J. Comput. Sci. 43(3) (2016)

35. Nikam, S.B., Agarwal, S.: Texture and wavelet-based spoof fingerprint detection for fingerprint biometric systems. In: Proc. International Conference on emerging Trends in Engineering and Technology (ICE-TET’), pp. 675–680 (2008)

36. Abhayankar, A., Schuckers, S.: Fingerprint liveliness detection using local ridge frequencies and multiresolution texture analysis techniques. In: Proc. International Conference on image Processing (ICIP), pp. 321–324 (2006)

37. Xie Z., et al.: Fingerprint liveness detection using gradient-based texture features. Signal Image Video Proc. 1–8 (2016)

38. Nogueira, R.F., de Alencar Lotufo, R., Machado, R.C.: Evaluating software-based fingerprint liveliness detection using convolutional networks and local binary patterns. In: Proc. International Workshop on Biometric Measurements and Systems for Security and Medical Applications (BIOMS), pp. 22–29 (2014)

39. Krizhevsky, A., Sutskever, I., Hinton, G.E.: Imagenet classification with deep convolutional neural networks. In: Advances in Neural Information Processing Systems, pp. 1097–1105 (2012)

40. Simonyan, K., Zisserman, A.: Very deep convolutional networks for large-scale image recognition (2014) arXiv preprint arXiv:1409.1556

41. Chugh, T., Jain, A.K.: Fingerprint presentation attack detection: Generalization and efficiency. In: Proc. International Conference on Biometrics (ICB), pp. 1–8.IEEE (2019)

42. Park, E. et al.: “IEEE Trans. on Information Forensics and Security (2019) Presentation attack detection using a tiny fully convolutional network

43. Zhang, Y. et al.: Light dense cnn for fingerprint liveness detection. IEEE Access. 8, 84141–84152 (2020)

44. Vedaldi, A., Fulkerson, B., Feat, V.L.: An open and portable library of computer vision algorithms (2008) http://www.vlfeat.org/

45. Csurka, G. et al.: Visual categorization with bags of keypoints. In: Proc. International Workshop on Statistical Learning in Computer Vision, pp. 1–22.ECCV (2004)

46. Dubey, R.K., Goh, J., Thing, V.: Fingerprint liveness detection from single image using low-level features and shape analysis. IEEE Trans. Inforn. Forensics Secur. 11(7), 1461–1475 (2016)

47. Silva, C., Bouwman, T., Frélicot, C.: An extended center-symmetric local binary pattern for background modeling and subtraction in videos. In: Proc. International Joint Conference on computer vision, Imaging and Computer Graphics Theory and Applications (VISAPP), Berlin (2015)

48. González-Soler, L.J., Gomez-Barrero, M., Busch, C.: Fisher vector encoding of dense-BSIF features for unknown face presentation attack detection. In: Proc. International Conference on the Special Interest Group on Biometrics (BIOSIG 2020), ser. LNI, GISeptember, pp. 1–11 (2020)

49. Martínez-Díaz, Y. et al.: Efficient video face recognition by using Fisher vector encoding of binary features. In: Proc. International Conference on Pattern Recognition, pp. 1436–1441.ICPR (2016)

50. Rosten, E., Drummond, T.: Machine learning for high-speed corner detection. In: Proc. International Conference on Computer Vision (ECCV), pp. 430–443 (2006)

51. Perronnin, F., Sánchez, J., Mensink, T.: Improving the Fisher kernel for large-scale image classification. In: Proc. International European Conference on Computer Vision (ECCV), pp. 143–156 (2010)

52. Jegou, H. et al.: Aggregating local image descriptors into compact codes. IEEE Trans. Pattern Anal. Mach. Intell. 34(9), 1704–1716 (2012)

53. Hsu, C., et al.: A Practical Guide to Support Vector Classification (2003)

54. Yamabay, D. et al.: LivDet 2011 fingerprint liveness detection competition 2011. In: 5th IAPR International Conference on Biometrics (ICB), pp. 208–215.IEEE (2012)

55. Ghiani, L., et al.: LivDet 2013 fingerprint liveness detection competition 2013. In: International Conference on Biometrics (ICB), pp. 1–6.IEEE (2013)

56. Mura, V. et al.: LivDet 2015 fingerprint liveness detection competition 2015. In: 7th International Conference on Biometrics: Theory, Applications and Systems (BTAS), pp. 1–6.IEEE (2015)

57. Mura, V. et al.: LivDet 2017 fingerprint liveness detection competition 2017. In: Proc. International Conference on Biometrics, (ICB), pp. 297–302 (2018)

58. Tabassi, E., NIST: Development of nfiq 2.0 (2015). https://www.nist.gov/services-resources/software/development-nfiq-20. Accessed 24 August 2020

59. Ghiani, L. et al.: Review of the fingerprint liveness detection (LivDet) competition series: 2009 to 2015. Image Vis. Comput. 58, 110–128 (2017)

How to cite this article: González-Soler L.J., Gomez-Barrero M., Kolberg J., Chang L., Pérez-Suárez A., Busch C. Local feature encoding for unknown presentation attack detection: An analysis of different local feature descriptors. IET Biome. 2021;1–18. https://doi.org/10.1049/bme2.12023