Fisher Vector Encoding of Dense-BSIF Features for Unknown Face Presentation Attack Detection

Lázaro J. González-Soler[1], Marta Gomez-Barrero[2], Christoph Busch[1]

Abstract: The task of determining whether a sample stems from a real subject (i.e, it is a bona fide presentation) or it comes from an artificial replica (i.e., it is an attack presentation) is a mandatory requirement for biometric capture devices, which has received a lot of attention in the recent past. Nowadays, most face Presentation Attack Detection (PAD) approaches have reported a good detection performance when they are evaluated on known Presentation Attack Instruments (PAIs) and acquisition conditions, in contrast to more challenging scenarios where unknown attacks are included in the evaluation. For those more realistic scenarios, the existing approaches are in many cases unable to detect unknown PAI species. In this work, we introduce a new feature space based on Fisher vectors, computed from compact Binarised Statistical Image Features (BSIF) histograms, which allows finding semantic feature subsets from known samples in order to enhance the detection of unknown attacks. This new representation, evaluated over three freely available facial databases, shows promising results in the top state-of-the-art: a BPCER100 under 17% together with a AUC over 98% can be achieved in the presence of unknown attacks.

Keywords: Presentation attack detection, probabilistic visual vocabulary, common feature space, unknown attacks, face.

1 Introduction

The deployment of biometric systems has increased over the last decades. In spite of their advantages, facial recognition systems are also vulnerable to Presentation Attacks (PAs): with the broad development experienced by social networks, an attacker can easily download a photo or video of a given person, thereby gaining access to several applications in which face recognition systems are commonly deployed. Moreover, the recent advances in creating synthetic videos, or deep fakes, also poses a serious threat [To20].

In order to address those concerns, several Software-based face Presentation Attack Detection (PAD) methods have been proposed. In general, many PAD approaches have reported a high detection performance for identifying Presentation Attack Instruments (PAIs) when both the attack type and acquisition conditions are known a priori (i.e., known attacks scenario). In contrast, a rather limited number of works have addressed so far a more realistic and challenging scenario where the PAI species in the test set remain unknown in the training set (i.e., unknown attacks). Back in 2013, de Freitas Pereira *et al.* [Fr13] already reported poor generalisation capabilities to unknown attacks of state-of-the-art face PAD methods based on local binary patterns (LBP) and support vector machines (SVMs). In

[1] dasec - Biometrics and Internet Security Research Group, Hochschule Darmstadt, Germany, {lazaro-janier.gonzalez-soler;christoph.busch}@h-da.de

[2] Hochschule Ansbach, Germany, marta.gomez-barrero@hs-ansbach.de

particular, the error rates increased by at least 100%. Motivated by these findings, Arashloo *et al.* [AKC17] experimented over several unknown attack scenarios and concluded that anomaly detection approaches trained only on bona fide data can reach a detection performance comparable to two-class classifiers. However, the results are reported only in terms of the area under the Receiving Operating Characteristic curve (AUC), thus lacking a proper quantitative analysis in line with the ISO/IEC 30107-3 standard on biometric PAD [IS17].

More recently, Nikisins *et al.* [Ni18] showed how a one-class Gaussian Mixture Model (GMM) can outperform two-class classifiers, depending on the PAI species included in the test set. The experimental evaluation reported an error rate increase with respect to the known scenario and two-class classifiers. Following the same anomaly detection paradigm, Xiong and AbdAlmageed studied in [XA18] the detection performance of one-class SVMs and autoencoders in combination with LBP descriptors for PAD purposes. In most of the scenarios tested, the detection rates increased with respect to common two-class classifiers. Finally, Liu *et al.* also analysed in [Li19] the performance of a Deep Tree Network (DTN) for facial PAD, which clusters the PAI species into semantic sub-groups. Over a new database comprising 13 PAI species, and following a leave-one-out testing protocol, an average D-EER of 16% is achieved, which is still above the state-of-the-art for known attacks.

To tackle those open issues with unknown attacks, we focus on a different approach which has already shown remarkable results in cross-sensor and unknown attacks scenarios for fingerprint PAD [Go19a]. Gonzalez-Soler *et al.* proposed in [Go19a] a combination of local feature descriptors and global feature encoding to model a new feature space in which the generalisation capabilities of the PAD module are enhanced. In fact, this approach achieved the best detection accuracy in the LivDet 2019 competition [Or19]. Whereas some keypoint based descriptors such as SIFT and SURF have shown to be an appropriate choice for fingerprint samples [Go19a, Go19b], in which minutiae can be regarded as landmarks within the image, we anticipate that for facial images the textural information is more relevant than the geometric information related to facial landmarks. Therefore, we propose a new face PAD approach, which encodes accurate and compact dense Binarized Statistical Image Features (dense-BSIF), extracted from local patches of the facial image, and projects them into a new feature space with Fisher vectors. By assuming that the unknown attacks share more texture, shape and appearance features with known PAIs than with those BP samples, this FV representation allows in turn a definition of semantic subgroups from known samples to tackle the aforementioned issues on PAD generalisation to unknown attacks. In order to validate the detection capabilities of the proposal, a thorough evaluation compliant with the ISO/IEC 30107-3 standard on biometric PAD [IS17] is also carried out over three well-established databases: CASIA Face Anti-Spoofing [Zh12], REPLAY-ATTACK [CAM12] and REPLAY-MOBILE [Co16].

The remainder of this paper is organised as follows. The proposed PAD method is presented in Sect. 2. Sect. 3 describes the experimental protocol and presents the results. Finally, conclusions and future work directions are presented in Sect. 4.

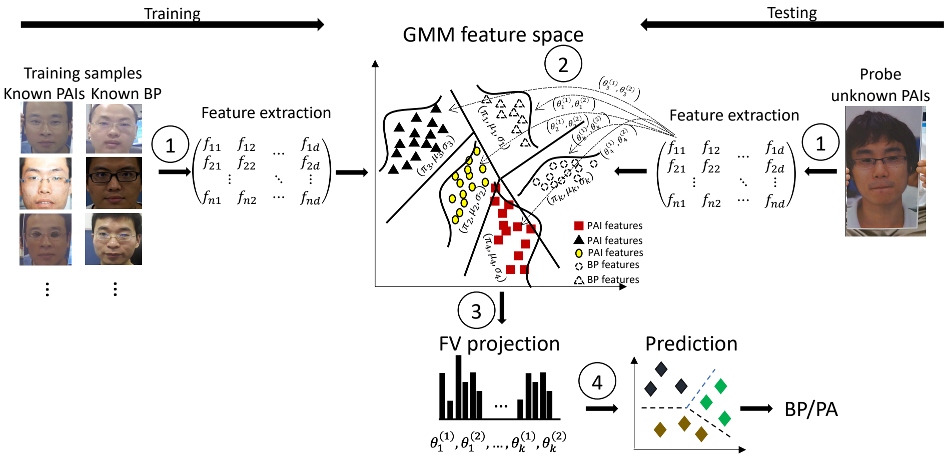

Training
Testing
GMM feature space
Training samples
Known PAIs Known BP
Feature extraction
1
$\begin{pmatrix} f_{11} & f_{12} & \cdots & f_{1d} \\ f_{21} & f_{22} & & f_{2d} \\ \vdots & & \ddots & \vdots \\ f_{n1} & f_{n2} & \cdots & f_{nd} \end{pmatrix}$
2
$(\theta_3^{(1)}, \theta_3^{(2)})$
$(\theta_1^{(1)}, \theta_1^{(2)})$
$(\theta_2^{(1)}, \theta_2^{(2)})$
$(\theta_k^{(1)}, \theta_k^{(2)})$
$(\theta_4^{(1)}, \theta_4^{(2)})$
(π_1, μ_1, σ_1)
(π_2, μ_2, σ_2)
(π_3, μ_3, σ_3)
(π_4, μ_4, σ_4)
(π_k, μ_k, σ_k)
🟥 PAI features
▲ PAI features
○ PAI features
◌ BP features
△ BP features
Probe
unknown PAIs
3
FV projection
$\theta_1^{(1)}, \theta_1^{(2)}, \ldots, \theta_k^{(1)}, \theta_k^{(2)}$
4
Prediction
BP/PA

Fig. 1: Face PAD approach overview which comprises three steps: *i*) BSIF descriptors are densely extracted; *ii*) the BSIF data distribution is subsequently learnt by training an unsupervised Gaussian Mixture Model (GMM) from known samples; *iii*) an unknown sample at hand is then encoded by computing the gradient among their BSIF components and parameters obtained by the GMM; and *iv*) the final features are finally classified using a linear SVM.

2 Proposed Method

We build upon the fingerprint PAD approach presented in [Go19a]. Fig. 1 shows an overview of the proposed PAD approach, which consists of four main steps: (1) dense-BSIF histograms are extracted from a face sample, which has been detected by the Viola and Jones method [VJ04]; (2) our new feature space is built by learning an unsupervised Gaussian Mixture Model (GMM) model from the aforementioned features; (3) the final descriptors are subsequently encoded by computing the differences of first- and second-order statistics with respect to the learned model parameters; and (4) a bona fide presentation (BP) or presentation attack (PA) decision is taken by a linear SVM.

2.1 Dense-BSIF Descriptors

Usually, PAIs include artefacts (e.g. acute edges in the cut eyes of the CASIA images [Zh12]) which can be successfully detected by the quantization of filtered features. BSIF [KR12] 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 $W = \{W_1, \ldots, W_N\}$ a set of linear filters of the same size as X. Then, we compute binarised responses b_n:

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

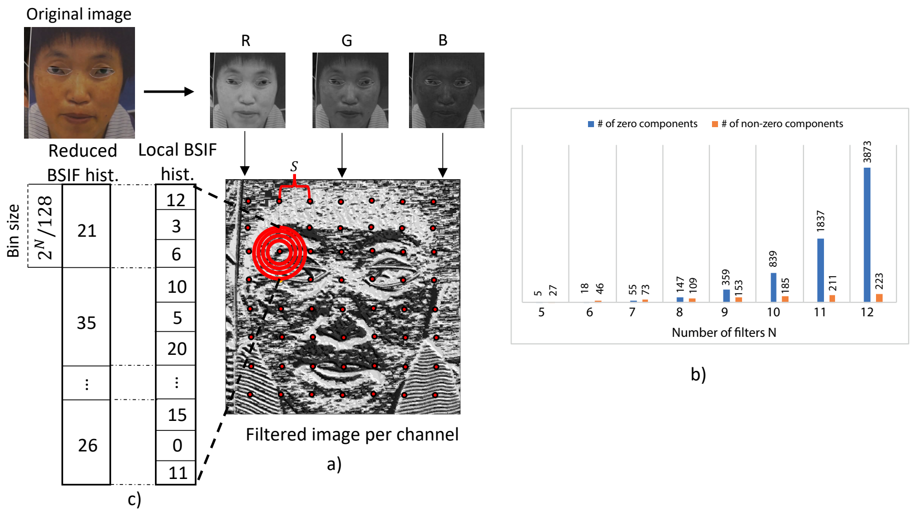


Fig. 2: BSIF feature extraction. a) BSIF histograms are densely computed at fixed points on a regular grid with a stride of S pixels, b) average number of zero and non-zero components of dense-BSIF histograms for different numbers of filters N, and c) a reduction example where a local BSIF histogram of size $2^N = 512$ is represented as a 128-component vector.

All the filter responses b_n are subsequently stacked to form a bit string $\mathbf{b}$ with size N for each pixel. Subsequently, $\mathbf{b}$ 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 [KR12].

Now, given that artefacts can be detected at any point in the image, not only at relevant facial landmarks, BSIF-histograms are densely extracted over a regular grid with a fixed stride S of 3. Furthermore, in order to capture local and global information of the artefacts produced in the fabrication of the PAIs, histograms are computed over four circular patches with different radii $r = \{4,6,8,10\}$, as depicted in Fig. 2a). Therefore, each point in the grid is represented by four dense-BSIF histograms.

In addition, we have observed that the histograms become sparse vectors as the number of linear filters N increases. Therefore, we computed the number of zero and non-zero components per number of filters over the CASIA Face Anti-Spoofing database [Zh12] in Fig. 2b), and noted that the number of non-zero components remains under 223 in all cases, having an average value of 128. We will thus represent each 2^N BSIF histogram as a 128-component vector by summing the elements for each sequential $2^N/128$ sub-set in the original histogram (see an example in Fig 2c)). This representation reduces the storage requirements down to 12.5% for N = 10 or 3.1% for N = 12.

2.2 Fisher Vector Encoding

The Fisher vector (FV) feature encoding approach transforms local features into an new feature space based on the parameters learnt by a generative model, known as visual vocabulary [PSM10]. This representation describes how the distribution of these local descriptors extracted from unknown PAIs differs from the known PAI distribution previously learned. In particular, a Gaussian Mixture Model (GMM) with 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 compute those semantic groups, the compact dense-BSIF descriptors (see Sect. 2.1) are decorrelated using Principal Component Analysis (PCA) [Je12], hence reducing their size to $d = 64$ components while retaining 98% of the variance. Then, the average first-order and second-order differences between the given decorrelated features and each semantic sub-group are computed, thereby obtaining a $2Kd$ dimensional vector. For the GMM computation, we selected $K = 1024$, since it allows representing a more complex structure from data while preserving low computational requirements. Therefore, each facial image is finally represented by a vector with size $2Kd = 2 \cdot 1024 \cdot 64 = 131072$.

3 Experimental Evaluation

3.1 Experimental Protocol

The experimental evaluation was conducted over three well-established databases for facial PAD, whose images were captured with different resolutions and several acquisition conditions: CASIA Face Anti-Spoofing [Zh12], REPLAY-ATTACK [CAM12], and REPLAY-MOBILE [Co16].

The experimental protocol aims to address the following goals: *i*) analyse the impact of different BSIF filter configurations in terms of number of filters and filter's size on the detection performance of our PAD approach, *ii*) study its detection performance for each RGB colour component under known and unknown attacks, and *iii*) benchmark the detection performance of our proposed PAD approach against the top state-of-the-art. Keeping these goals in mind, we defined two different scenarios:

- *Known-attacks*: which includes an analysis of all PAI species. In all cases, PAI species for testing are included in the training set, as described in [Zh12].
- *Unknown-attacks*, in which the PAI species used for testing are not incorporated in the training set. In particular, we consider using the leave-one-out testing protocol explained in [AKC17], in which a PAI is only evaluated whilst the remaining PAI species are employed for training.

Finally, the experimental evaluation is conducted in compliance with the international metrics of ISO/IEC 30107-3 [IS17]: *i*) Attack Presentation Classification Error Rate (APCER)

Tab. 1: Benchmark in terms of D-EER(%) of our PAD approach per colour component and the whole RGB colour space against the top state-of-the-art methods.

Method	**Warped**	**Cut**	**Video**	**Overall**
BSIF + SVM [RB14]	-	-	-	10.21
MBSIF-TOP [AKC15]	1.40	10.10	4.30	7.20
Texture fusion [BKH18]	-	-	-	4.60
ResNet-15 on 3D [Gu19]	-	-	-	2.22
shallowCNN-LE [QDN19]	-	-	-	4.00
SPMT + SSD [So19]	0.35	0.20	0.03	0.04
Hybrid residual DL [MH19]	-	-	-	0.02
Proposed Method (R)	1.42 ± 1.04	2.20 ± 0.83	0.28 ± 0.56	2.92 ± 0.93
Proposed Method (G)	1.44 ± 1.11	2.22 ± 0.79	0.50 ± 0.72	2.52 ± 1.18
Proposed Method (B)	1.59 ± 1.01	2.78 ± 1.27	0.59 ± 0.77	2.53 ± 0.99
Proposed Method (RGB)	1.20 ± 0.77	1.74 ± 0.75	0.30 ± 0.54	1.79 ± 0.82

which is defined as the proportion of attack presentations wrongly classified as bona fide presentations, and *ii*) Bona Fide Presentation Classification Error Rate (BPCER) which is the proportion of bona fide presentations missclassified as attack presentations. We therefore report: *i*) the Detection Error Trade-off (DET) curves between both APCER and BPCER, *ii*) the BPCER values for several security thresholds (BPCER10, BPCER20 and BPCER100), and *iii*) the Detection Equal Error Rate (D-EER), which are defined as the error rate value at the operating point where APCER = BPCER.

3.2 Results and Discussion

3.2.1 Known Attacks

First, we need to find the optimal configuration of our proposed method in terms of the filter size l, the number of BSIF filters N, and the best performing RGB component. To that end, we compute error rates for each of sixty filter configuration and report in Table. 1 the mean and standard deviation (std) of the D-EER achieved by each particular RGB component and their fusion (i.e., RGB) over the Attack test and Overall test protocol in the CASIA Face Anti-Spoofing database [Zh12]. As it could be expected, the entire RGB space reports on average the best detection performance, since it fuses the information of the three channels. In addition, it reports for N = 11 filters of size l = 11 a minimum D-EER of 0.0% for warped, 1.11% for cut photo attacks, 0.0% for video replay attacks, and 0.37% for the overall test, which achieve the top state-of-the-art.

On the other hand, it may also be observed that among the three individual RGB colour components, the R channel appears to be the one which contains the most discriminative features for the PAD task. Specifically, it achieves for $N = 9$ filters of size l = 6, a D-EER of 0.0% for warped photo attacks, 2.22% for cut photo attacks, 0.0% for video replay attacks, Tab. 2: D-EER(%) values under the *Unknown-attacks* protocol for RGB and benchmark, in terms of AUC(%), for the best unknown attack setting (i.e., $N = 10$ filters of size $l = 9$), against the top state-of-the-art approaches.

	CASIA			**REPLAY-ATTACK**			**REPLAY-MOBILE**		
	Cut	Warped	Video	Digital	Printed	Video	Digital	Printed	Video
OC-SVM$_R GB$+BSIF [AKC17]	60.7	95.9	70.7	88.1	73.7	84.3	-	-	-
NN+LBP [XA18]	88.4	79.9	94.2	95.2	78.9	99.8	-	-	-
DTL [Li19]	97.3	97.5	90.0	99.9	99.6	99.9	-	-	-
our proposal (AUC)	**99.6**	**97.9**	**99.9**	**100**	**99.9**	**100**	**100**	**100**	**100**
our proposal (D-EER)	4.11 ± 1.99	6.15 ± 2.42	1.37 ± 1.60	0.00 ± 0.00	1.35 ± 1.73	0.00 ± 0.00	0.00 ± 0.00	0.34 ± 0.63	0.02 ± 0.12

and 0.37% for the overall test, which are almost the same minimum D-EER reported by the entire RGB colour space.

In order to validate the detection performance of our PAD approach under different RGB configurations, we select the non-parametric Mann-Whitney test with a 95% of confidence and verify the statistical significance of error rates reported by each RGB component. To that end, we define as null and alternative hypothesis:

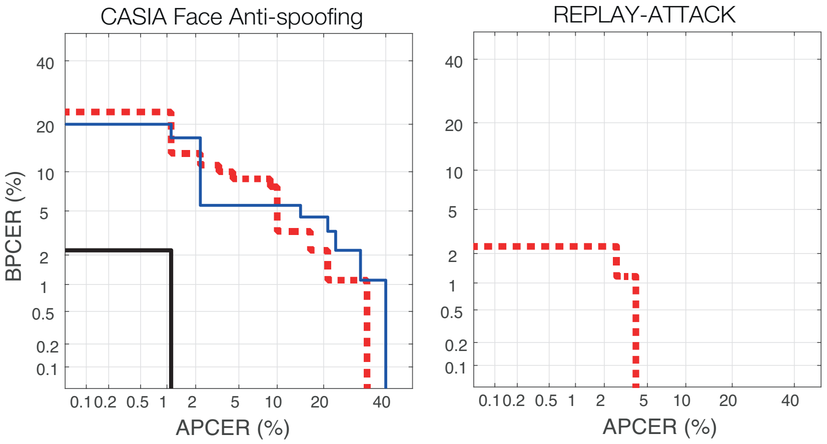
- H_0: two colour components provide the same discriminative information for PAD.
- H_1: two colour components do not provide the same discriminative information for PAD.

Then, an all-against-all comparison per database is performed. As a result of this test, we do confirm that error rates for the three RGB colour components stem from the same population, thereby showing the same discriminative information for facial PAD. In contrast, results for the entire RGB colour space claim to be statistically better than each particular component. Therefore, we do confirm that even if each individual RGB component is correlated with each other, the entire RGB colour space includes some additional information which allows learning more discriminative features for facial PAD.

3.2.2 Unknown Attacks

As it was mentioned in Sect. 1, one of the goals of our work is to successfully identify unknown PAIs. To that end, a set of experiments is carried out over the three selected databases following the leave-one-out protocol described in [AKC17]: two PAI species are included in the training set, and the last one in the test set. Table 2 reports the corresponding D-EER values, a benchmark against the top state-of-the-art PAD approaches and the complete DET curves are depicted in Fig. 3.

Taking a look at Table 2, we can observe how error rates for each particular unknown PAIs with respect to the corresponding known attack are multiplied by a factor of 2.36% for cut photo attacks, 4.57% for video replay attacks and 5.13% for warped photo attacks, respectively, for the CASIA database. However, for a fixed filter configuration, it should be noted



Train vs test	BPCER10	BPCER20	BPCER100
— warped + video vs cut	5.56	5.56	16.7
▪▪▪ video + cut vs warped	3.33	8.89	13.3
▬ cut + warped vs video	0.00	0.00	0,00

Train vs test	BPCER10	BPCER20	BPCER100
— printed + video vs digital	0.00	0.00	0,00
▪▪▪ digital + video vs printed	0.00	0.00	2.50
▬ printed + digital vs video	0.00	0.00	0,00

Fig. 3: *Unknown-attacks* DET curves and BPCER(%) values over the leave-one-out protocol for the CASIA and REPLAY-ATTACK databases. The REPLAY-MOBILE database reports a BPCER = 0.0% for any APCER.

that our proposed approach outperforms, in terms of AUC, the top state-of-the-art for all attack types. Since we lack a proper quantitative analysis of the top state-of-the-art methods in compliance with the ISO/IEC 30107-3 standard on biometric PAD [IS17], we can only establish a benchmark in terms of AUC. For the REPLAY's databases, a higher detection performance outperforming the top-state-of-the-art techniques can be observed: an AUC of almost 100% for the entire set of attacks confirms the soundness of our proposed method to identify PAIs stemming from this challenging scenario.

In addition, Fig. 3 confirms the detection performance showed by our approach: a low BPCER100 of 0.0%, 0.0%, 13.3% and 16.7% are achieved by video replay, digital, cut photo and warped photo attacks, respectively for a high security threshold (i.e., APCER of 1%). This in turn yields, in this challenging scenario, a secure (only one in 100 attacks are not detected) and convenient (zero to seventeen in 100 bona fide presentation attempts are rejected) system.

Finally, a t-SNE visualisation in Fig. 4 of BP and PA samples are in the CASIA database confirms the aforementioned hypotheses, which state that the PAIs share more texture, shape and appearance features with known PAIs than with those BP samples. Whereas the FV representations of attack presentations (blue, red and yellow) are separated of the bona fide presentations (green spots), they are close to each other. However, we can also observe that some PAIs, such as warped (yellow) and cut photo attacks (blue), still overlap with

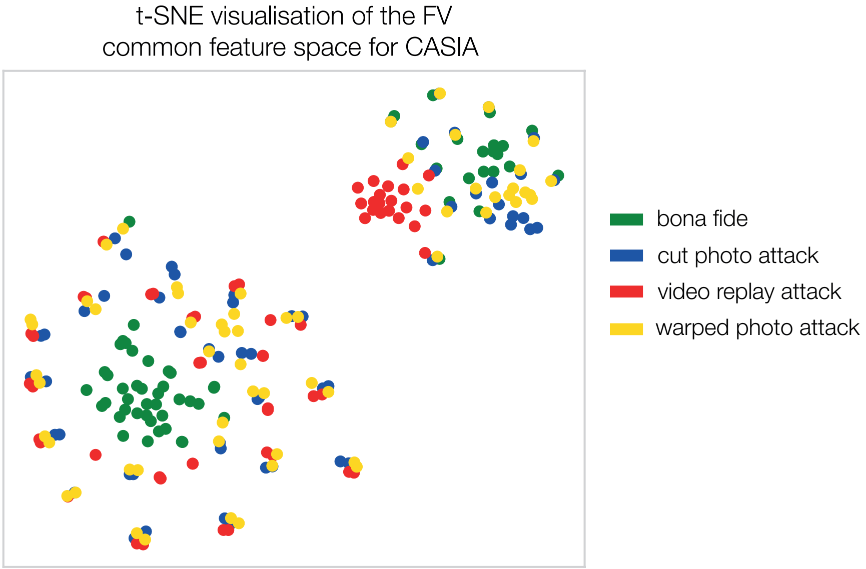


Fig. 4: t-SNE visualisation for BP vs. PA samples in the CASIA Face Anti-spoofing database.

BP samples, thereby indicating that the data distribution learned by a GMM model using the BSIF features needs to be improved in order to get a better detection performance.

4 Conclusions

In this work, a new face PAD approach to generalise to unknown attacks was proposed. In essence, it projects compact dense-BSIF descriptors into a new feature space, which allows discovering semantic feature sub-groups from known samples in order to improve the PAD generalisation capability. In addition, a new strategy for computing compact dense-BSIF histograms was presented, which can be applied to any other texture recognition application. In more details, a reduction down to 95% in the feature vector length can be achieved with no significant impact on the recognition accuracy but strongly reducing the time required for PAD analysis. The experimental evaluation over three freely available databases confirmed the soundness of our proposal for detecting both known and unknown PAIs. Specifically, experimental results indicated the statistical advantage of the entire RGB colour space with respect to each of its particular components, thereby resulting in a minimum D-EER of 0.37% for known attack detection. Finally, BPCER100 in a range of 0.0% to 17% for unknown attack detection, which outperform the top state-of-the-art, showed that our PAD approach is able to yield a secure and convenient system under that challenging scenario. As future work, we plan to evaluate our proposal on larger databases for other colour spaces, which have shown to be superior in terms of detection performance. In addition, a more thorough analysis on larger databases including a higher number of different PAI species will be carried out.

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 Hessen State Ministry for Higher Education, Research and the Arts within their joint support of the National Research Center for Applied Cybersecurity ATHENE.

References

[AKC15] Arashloo, S. R.; Kittler, J.; Christmas, W.: Face spoofing detection based on multiple descriptor fusion using multiscale dynamic binarized statistical image features. IEEE Trans. on Information Forensics and Security, 10(11):2396–2407, 2015.

[AKC17] Arashloo, S. R.; Kittler, J.; Christmas, W.: An anomaly detection approach to face spoofing detection: A new formulation and evaluation protocol. IEEE Access, 5:13868–13882, 2017.

[BKH18] Boulkenafet, Z.; Komulainen, J.; Hadid, A.: On the generalization of color texture-based face anti-spoofing. Image and Vision Computing, 77:1–9, 2018.

[CAM12] Chingovska, I.; Anjos, A.; Marcel, S.: On the Effectiveness of Local Binary Patterns in Face Anti-spoofing. 2012.

[Co16] Costa-Pazo, A.; Bhattacharjee, S.; Vazquez-Fernandez, E.; Marcel, S.: The REPLAY-MOBILE Face Presentation-Attack Database. In: Proc. Intl. Conf. on Biometrics Special Interests Group (BIOSIG). 2016.

[Fr13] de Freitas Pereira, T.; Anjos, A.; Martino, J. De; Marcel, S.: Can face anti-spoofing countermeasures work in a real world scenario? In: Proc. Int. Conf. on Biometrics (ICB). pp. 1–8, 2013.

[Go19a] González-Soler, L. J.; Gomez-Barrero, M.; Chang, L.; Pérez-Suárez, A.; Busch, C.: Fingerprint Presentation Attack Detection Based on Local Features Encoding for Unknown Attacks. arXiv preprint arXiv:1908.10163, 2019.

[Go19b] González-Soler, L. J.; Gomez-Barrero, M.; Chang, L.; Pérez-Suárez, A.; Busch, C.: On the Impact of Different Fabrication Materials on Fingerprint Presentation Attack Detection. In: Proc. Intl. Conf. on Biometrics (ICB). 2019.

[Gu19] Guo, J.; Zhu, X.; Xiao, J.; Lei, Z.; Wan, G.; Li, S. Z.: Improving Face Anti-Spoofing by 3D Virtual Synthesis. In: Proc. Intl. Conf. on Biometrics (ICB). 2019.

[IS17] ISO/IEC JTC1 SC37 Biometrics: . ISO/IEC FDIS 30107-3. Information Technology - Biometric presentation attack detection - Part 3: Testing and Reporting. International Organization for Standardization, 2017.

[Je12] Jegou, H.; Perronnin, F.; Douze, M.; Sánchez, J.; Perez, P.; Schmid, C.: Aggregating local image descriptors into compact codes. IEEE Trans. on Pattern Analysis and Machine Intelligence, 34(9):1704–1716, 2012.

[KR12] Kannala, J.; Rahtu, E.: BSIF: Binarized statistical image features. In: 2012 21st Intl. Conf. on Pattern Recognition (ICPR). pp. 1363–1366, 2012.

[Li19] Liu, Yaojie; Stehouwer, Joel; Jourabloo, Amin; Liu, Xiaoming: Deep Tree Learning for Zero-shot Face Anti-Spoofing. In: Proc. Conf. on Computer Vision and Pattern Recognition. pp. 4680–4689, 2019.

[MH19] Muhammad, U.; Hadid, A.: Face Anti-spoofing using Hybrid Residual Learning Framework. In: Proc. Intl. Conf. on Biometrics (ICB). 2019.

[Ni18] Nikisins, O.; Mohammadi, A.; Anjos, A.; Marcel, S.: On effectiveness of anomaly detection approaches against unseen presentation attacks in face anti-spoofing. In: Proc. Int. Conf. on Biometrics (ICB). pp. 75–81, 2018.

[Or19] Orrù, G.; Casula, R.; Tuveri, P.; Bazzoni, C.; Dessalvi, G.; Micheletto, M.; Ghiani, L.; Marcialis, G. L.: Livdet in action-fingerprint liveness detection competition 2019. In: Proc. Intl. Conf. on Biometrics (ICB). IEEE, pp. 1–6, 2019.

[PSM10] Perronnin, F.; Sánchez, J.; Mensink, T.: Improving the fisher kernel for large-scale image classification. In: Proc. European Conf. on Computer Vision (ECCV). pp. 143–156, 2010.

[QDN19] Qu, X.; Dong, J.; Niu, S.: shallowCNN-LE: A shallow CNN with Laplacian Embedding for face anti-spoofing. In: Intl. Conf. on Automatic Face & Gesture Recognition. pp. 1–8, 2019.

[RB14] Raghavendra, R.; Busch, C.: Presentation attack detection algorithm for face and iris biometrics. In: Proc. European Signal Processing Conf. (EUSIPCO). pp. 1387–1391, 2014.

[So19] Song, X.; Zhao, X.; Fang, L.; Lin, T.: Discriminative representation combinations for accurate face spoofing detection. Pattern Recognition, 85:220–231, 2019.

[To20] Tolosana, R.; Vera-Rodriguez, R.; Fierrez, J.; Morales, A.; Ortega-Garcia, J.: DeepFakes and Beyond: A Survey of Face Manipulation and Fake Detection. arXiv preprint arXiv:2001.00179, 2020.

[VJ04] Viola, P.; Jones, M. J.: Robust real-time face detection. Intl. Journal of Computer Vision, 57(2):137–154, 2004.

[XA18] Xiong, F.; AbdAlmageed, W.: Unknown presentation attack detection with face RGB images. In: Proc. Int. Conf. on Biometrics Theory, Applications and Systems (BTAS). pp. 1–9, 2018.

[Zh12] Zhang, Z.; Yan, J.; Liu, S.; Lei, Z.; Yi, D.; Li, S. Z: A face antispoofing database with diverse attacks. In: Proc. Intl. Conf. on Biometrics (ICB). pp. 26–31, 2012.