Fingerprint Presentation Attack Detection Method Based on a Bag-of-Words Approach

Lázaro Janier González-Soler[1(✉)], Leonardo Chang[3], José Hernández-Palancar[1], Airel Pérez-Suárez[1], and Marta Gomez-Barrero[2]

[1] Advanced Technologies Application Center (CENATAV), 7a # 21406 e/ 214 y 216, Rpto. Siboney, Playa, 12200, Havana, Cuba
{jsoler,jhernandez,asuarez}@cenatav.co.cu
[2] da/sec - Biometrics and Internet Security Research Group, Hochschule Darmstadt, Darmstadt, Germany
marta.gomez-barrero@h-da.de
[3] Tecnologico de Monterrey, Campus Estado de México, Carretera al Lago de Guadalupe Km. 3.5, 52926 Atizapán de Zaragoza, Estado de México, Mexico
leonardochang36@gmail.com

Abstract. Fingerprint-based biometric systems are not entirely secure due to their vulnerability to presentation attacks. In this paper, we propose a new presentation attack method based on a Bag-of-Words approach, which by combining local and global information of fingerprint can correctly identify bona fine presentations from presentation attacks. The experimental evaluation of our proposal, over the well-known LivDet 2011 dataset, showed an Average Classification Error of 4.73%, outperforming the state of the art.

1 Introduction

Fingerprint-based biometric systems are commonly used in several contexts since they provide higher accuracy and in many cases, they increase user convenience with respect to traditional credential-based systems. Despite being very popular, fingerprint-based biometric systems are vulnerable to presentation attacks, in which synthetic artifacts are presented to the biometric sensor. The task of determining if a fingerprint comes from a live subject (i.e., it is a bona fide presentation) or it comes from an artificial fingertip (i.e., it is a presentation attack) is a critical issue that has received a considerable attention in the recent years.

In this context, several software-based methods have been reported in literature for Presentation Attack Detection (PAD) [1,2]. These methods focus on skin properties, which can be classified as dynamic (e.g., skin distortion and perspiration) or static (e.g., texture, ridge frequency and ridge-valley structure). In contrast to hardware-based methods, these approaches are a less expensive alternative since they do not require a special hardware. In addition, most of the software-based methods use a single image and do not employ additional invasive biometric measurements. However, they can be more challenging, as

M. Mendoza and S. Velastín (Eds.): CIARP 2017, LNCS 10657, pp. 263–271, 2018.
https://doi.org/10.1007/978-3-319-75193-1_32

they require the extraction of discriminative features to distinguish a bona fide presentation from a presentation attack.

Taking into account that real and fake fingerprint images exhibit different appearance characteristics, many texture descriptors have been used for discriminating presentation attacks from bona fide presentations (e.g., LPQ [3], MBLTP [4], WLD [5], BSIF [6]), and combinations of individual algorithms (e.g., SURF-PHOG-Gabor [7]). The main drawback of the texture-based methods is that they depend both on the material used for building the artificial fingerprint and the sensor used for acquiring the fingerprint images. More specifically, different materials utilized for building artificial fingerprint replicas exhibit different ridge structures, which can be identified based on discriminative multi-scale appearance description methods.

An appearance-based technique that has shown remarkable results in object classification tasks is the Bag of Words (BoW) approach [8]. In this paper, we propose a new PAD method based on the BoW algorithm, which computes local descriptors at fixed points on a regular grid. In our methodology, several scaling factors are studied in order to represent both local and global information of a fingerprint, thereby allowing to distinguish a bona fide presentation from a presentation attack.

In order to evaluate the performance of the proposal and to allow the reproducibility of the presented results, several experiments were conducted on the well-known LivDet 2011 dataset. The proposed method achieved an Average Classification Error (ACE) of 4.73%, significantly outperforming the top state-of-the-art results. The performance of the proposal was also evaluated using the recent ISO standard evaluation metrics specifically tailored for PAD [9].

The remainder of this paper is organized as follows: in Sect. 2, we describe the proposed method. The experimental evaluation, comparing the performance of our proposal against top state-of-the-art techniques, is presented in Sect. 3. Finally, conclusions and future work directions are presented in Sect. 4.

2 Proposed Method

Bag-of-words approaches have been largely used for image appearance description and they have obtained remarkable results in object categorization tasks. In this paper, we propose a method for detecting fingerprint presentation attacks which is based on a BoW approach. Figure 1 shows an overview of the proposed method, which pipeline is composed of the following three steps: (i) extraction of Scale Invariant Feature Transform (SIFT) descriptors, (ii) encoding of the SIFT features with a spatial histogram of visual words, and (iii) classification of the image descriptor by a Support Vector Machine (SVM) through a feature map.

For local features extraction, we follow the Pyramid Histogram Of visual Words (PHOW) approach of [10], in which SIFT descriptors are computed densely at points on a regular grid with a fixed spacing (e.g., 5 pixels). In order to allow scale variation between fingerprints, descriptors are computed over four circular patches with different radii σ. Therefore, each point in the grid is represented by four SIFT descriptors. To efficiently compute the descriptors, we use the implementation provided in [11], which delivers a speed-up of up to 60x by

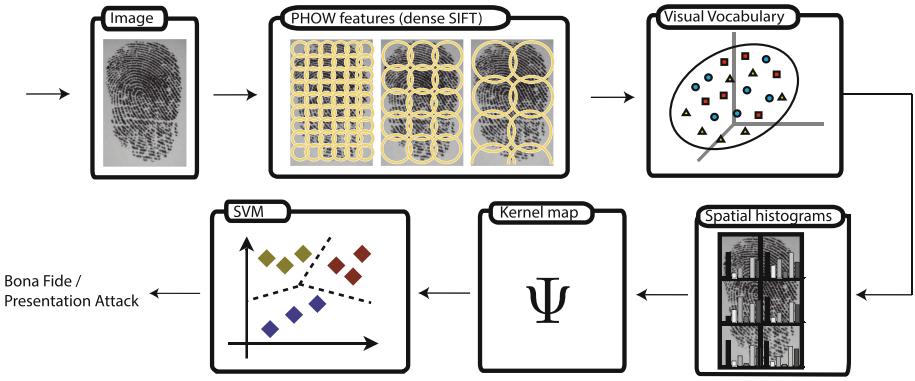

Fig. 1. Method overview. First, dense SIFT descriptors are computed at different scales. Then this features are encoded using a previously learned visual vocabulary. Finally, the fingerprint descriptor is classified using SVM through a feature map.

exploiting the uniform sampling and overlapping between descriptors, and by implementing linear interpolation with integral image convolution.

For image encoding, spatial histograms are used in order to incorporate spatial relationship between patches. In this context, the fingerprint image is partitioned into increasingly fine sub-regions and the dense extracted features inside each sub-region are vector quantized into K visual words using a k-means clustering. Once the BoW has been built, histograms inside each sub-region are computed and they are then stacked for describing the fingerprint image.

Finally, for bona fide or presentation attack classification, a homogeneous kernel map is used to transform a χ^2 SVM into a linear one [12], providing fast computation with high accuracy.

3 Experimental Evaluation

In this section, the results of several experiments for evaluating the performance of the proposed method are presented. The objectives of the experimental evaluation are threefold, namely: (i) analyse the impact of the different parameters on the method performance, (ii) compare the performance of our proposal against the top state-of-the-art approaches and (iii) analyse the computational time of the proposed algorithm for different parameter configurations.

3.1 Experimental Protocol

The experiments were conducted on the publicly available database provided by LivDet 2011 [13]. The database contains 16,000 images acquired with four different sensors, namely, Biometrika FX2000, Digital 4000B, Italdata ET10 and Sagem MSO300. For each sensor, 2,000 images from presentation attacks and 2,000 images from bona fide presentations were captured. Presentation attacks

were built using five different materials: gelatin, eco flex, latex, silgum and wood glue [14].

The proposed method was implemented in Matlab using the VLFeat library [11] for SIFT computation, k-means and SVM. All the experiments were performed on a PC with an Intel i5-3470 processor at 3.2 GHz, 8 GB RAM.

In all the experiments we followed the LivDet 2011 evaluation protocol, where half of the database is used for training and the other half for testing. In order to allow a fair comparison with the available literature, we reported the ACE, which is the standard metric in LivDet competitions

$$\text{ACE} = \frac{\text{FSAR} + \text{FLRR}}{2}, \tag{1}$$

where False Spoofing Acceptance Rate (FSAR) is the percentage of misclassified spoof fingerprints (i.e., presentation attacks) and False Live Rejection Rate (FLRR) is the percentage of misclassified live fingerprints (i.e., bona fide presentations).

In addition, we also reported the accuracy our proposal attains in terms of Attack Presentation Classification Error Rate (APCER) and Bona Fide Presentation Classification Error Rate (BPCER), which are recent ISO standard metrics used for evaluating PAD schemes [9].

3.2 Experimental Results

In the first set of experiments, we study the impact of the different key parameters; the visual vocabulary sizes (K) and the scale values (σ), on the performance and efficiency of our proposal. In particular, we have chosen the following ranges: $K = \{256, 512, 1024, 2048, 4096, 8192\}$ and $\sigma = \{1, 1.40, 1.96, 2.74, 3.84\}$. For determining the best performance of the PAD method, all combinations of K and σ are evaluated, hence resulting in 30 different experiments.

Figure 2 shows the impact on the ACE of scale variation for the different visual vocabulary sizes, for each sensor in the dataset. As it can be observed, most of the curves have a stationary behaviour (Biometrika, Digital and Sagem sensors) with the exception of the ItalData sensor, which shows a more random trend for different scales. Both trends hence indicate a lack of a significant impact of σ on the overall accuracy of the PAD scheme. Therefore, for applications where computational cost is a critical issue, any σ value can be used without considerably affecting the system accuracy.

In Fig. 3, we show the impact on the classification accuracy of the different vocabulary sizes (K) at different scale values (σ) for each sensor in the LivDet 2011 dataset. It can be appreciated that the size of the vocabulary has a significant impact on the classification accuracy. Figure 3 shows that again the Biometrika, Digital and Sagem sensors exhibit a similar behaviour, where the best results are obtained for $K = 1024$. On the other hand, for the ItalData sensor, the best results are obtained for the biggest vocabularies, $K = 8192$.

In the second experiment, we computed the ACE achieved by our proposal and we compared it with those yielded by the top state-of-the-art methods; these

results are showed in Table 1. As it can be observed, several configurations of our proposal achieve state-of-the-art results for the LivDet 2011. In particular, the best reported result presented in [14], based on deep learning, is outperformed by 14%. Compared to other appearance-based methods, our proposal outperforms the best two results by 19% [7] and 21% [15] of ACE, respectively.

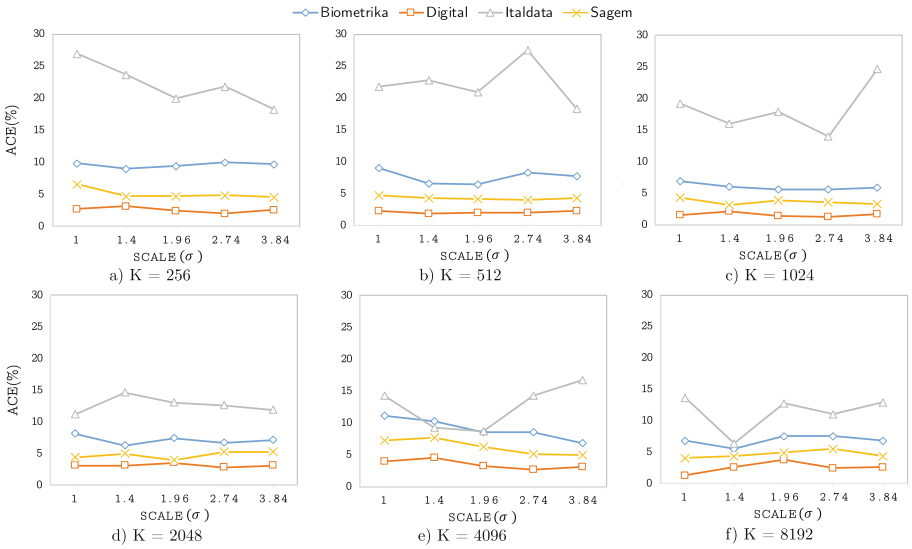

Fig. 2. ACE per sensor on several scales (σ) for different visual vocabulary sizes (K).

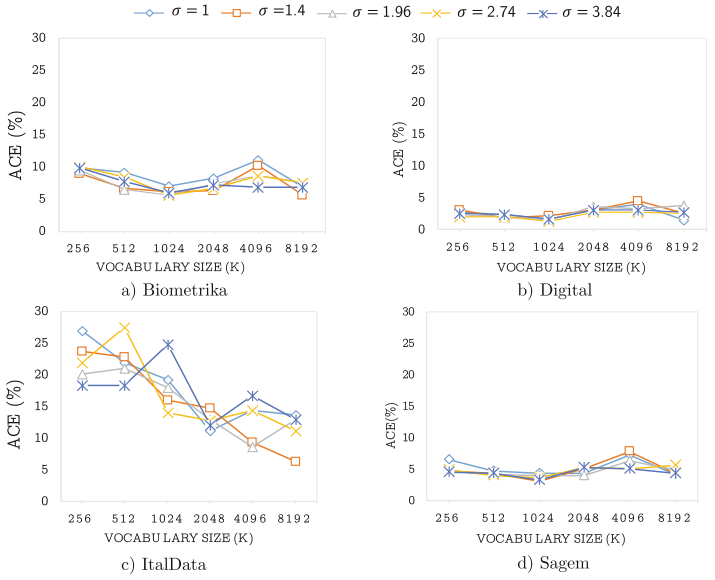

Fig. 3. ACE per scale for each sensor using different sizes of visual vocabulary.

Table 1. Comparison of ACE. For each sensor, the lowest value is highlighted in bold.

Methods	Biometrika	Digital	ItalData	Sagem	Avg. ACE
Our method (K = 1024, σ = 1.40)	6.05	2.15	16.05	**3.15**	6.85
Our method (K = 1024, σ = 2.74)	5.65	**1.25**	13.95	3.65	6.13
Our method (K = 2048, σ = 1.00)	8.15	3.15	11.15	4.35	6.70
Our method (K = 4096, σ = 1.96)	8.50	3.30	8.65	6.30	6.69
Our method (K = 8192, σ = 1.4)	**5.60**	2.60	6.35	4.35	**4.73**
Our method (K = 8192, σ = 3.84)	7.00	2.50	11.40	4.40	6.33
SURF+PHOG+Gabor 2016 [7]	7.89	6.25	8.10	5.36	6.90
Xia et al. method 2016 [15]	6.70	4.75	11.75	3.34	6.64
Wavelet + LBP 2016 [16]	8.20	7.20	11.85	7.86	8.78
Jiang and Liu 2015 [17]	8.45	15.35	14.85	5.26	10.98
Johnson and Schuckers 2014 [18]	18.4	7.80	15.20	6.70	12.03
Nogueira et al. 2014 [14]	9.90	1.90	**5.09**	7.86	6.19
MSLBP 2014 [19]	7.30	2.50	14.80	5.30	7.48
WLD 2013 [5]	7.20	8.00	12.65	3.66	7.87
BSIF 2013 [6]	6.80	3.55	13.65	4.86	7.22
MBLTP 2013 [4]	10.00	6.90	16.30	5.90	9.77
LPQ 2012 [3]	10.40	8.00	13.20	5.30	9.23

Finally, we evaluate the performance of our proposal using recent metrics for PAD defined within the ISO/IEC FDIS 30107-3 [9]. The proportion of bona fide presentations incorrectly classified as presentation attacks for a fixed APCER of 10% (BPCER10) and 5% (BPCER20) are also reported. Figure 4 shows the Detection Error Trade-Off Curve (DET) as well as the BPCER10 and BPCER20 values for the four LivDet 2011 sensors.

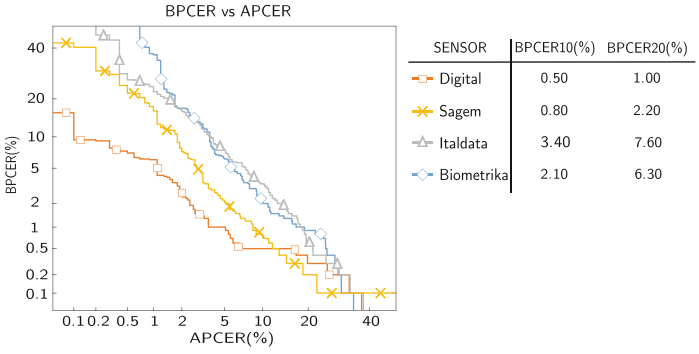

Fig. 4. DET curves, BPCER10 and BPCER20 values for the four sensors.

As it can be observed from Fig. 4, for smaller values of APCER (higher security thresholds), Digital sensor reports smallest values of BPCER. Specifically, it shows a BPCER value (BPCER10 = 0.5%), which is 7 times smaller than the one reported by Italdata sensor (BPCER10 = 3.4%). These results show that our proposed method is suitable for applications demanding very high security, in which presentation attacks may be frequently employed.

In the third experiment, we study the computational efficiency of the proposed algorithm for different parameters configurations. Table 2 shows the average performance of the proposal over different vocabulary sizes (K) at different scale values (σ). As it can be observed, different K and σ values have a slight impact in the average computational efficiency of the proposed method. Moreover, these efficiency results indicate that higher vocabulary sizes (K) and higher scale values (σ) worsen the computational efficiency of the proposal, at the same time, its accuracy is improved. On the other hand, an opposite effect occurs when K is close to 256. It should be noted that, in all cases, the efficiency value reported for each parameter combination is always below 810 ms; therefore, we can conclude that our proposal can be utilized in any real-time PAD system.

Table 2. Performance in milliseconds of our method for different configuration.

	$\sigma = 1$	$\sigma = 1.40$	$\sigma = 1.96$	$\sigma = 2.74$	$\sigma = 3.84$
K = 256	630.4	690.9	687.7	695.5	734.3
K = 512	642.3	657.9	672.9	697.7	732.1
K = 1024	673.3	680.7	700.3	722.1	753.3
K = 2048	677.5	706.2	721.4	729.0	758.2
K = 4096	692.3	707.7	713.8	753.5	805.9
K = 8192	700.6	742.6	752.6	774.3	807.8

4 Conclusions

In this paper, we proposed a new PAD method based on a BoW approach. The experimental evaluation conducted over the publicly available database LivDet 2011 assessed the performance of our proposal with respect to the top state-of-the-art methods. Experiments focused on parameter optimization showed that the scaling factor σ does not have a significant impact on the overall accuracy of the PAD, whilst the size of vocabulary (K) does. The proposal achieved the best result at $K = 8192$ and $\sigma = 1.4$, decrease the sate-of-the-art by 14%. This result shows the high capacity of the representation based on BoW in the task of PAD. Finally, the computational efficiency evaluation showed that our proposal is suitable for real-time applications, where the overall process took between 600 and 800 ms.

As future work, in order to improve the distinctiveness of the proposed descriptor, we plan to tackle some of the limitations of the histogram-based

representation by using VLAD and Fisher Vector representations which are known to be much more powerful, in terms of representation, than the plain histogram.

Acknowledgement. This work was partly supported by the German Federal Ministry of Education and Research (BMBF) as well as by the Hessen State Ministry for Higher Education, Research and the Arts (HMWK) within the Center for Research in Security and Privacy (CRISP).

References

1. Marasco, E., Ross, A.: A survey on antispoofing schemes for fingerprint recognition systems. ACM Comput. Surv. (CSUR) **47**(2), 28 (2015)
2. Sousedik, C., Busch, C.: Presentation attack detection methods for fingerprint recognition systems: a survey. Iet Biometrics **3**(4), 219–233 (2014)
3. Ghiani, L., Marcialis, G.L., Roli, F.: Fingerprint liveness detection by local phase quantization. In: ICPR 2012, pp. 537–540. IEEE (2012)
4. Jia, X., Yang, X., Zang, Y., Zhang, N., Dai, R., Tian, J., Zhao, J.: Multi-scale block local ternary patterns for fingerprints vitality detection. In: ICB 2013, pp. 1–6. IEEE (2013)
5. Gragnaniello, D., Poggi, G., Sansone, C., Verdoliva, L.: Fingerprint liveness detection based on Weber local image descriptor. In: 2013 IEEE Workshop on BIOMS, pp. 46–50. IEEE (2013)
6. Ghiani, L., Hadid, A., Marcialis, G.L., Roli, F.: Fingerprint liveness detection using binarized statistical image features. In: BTAS, pp. 1–6. IEEE (2013)
7. Dubey, R.K., Goh, J., Thing, V.L.: Fingerprint liveness detection from single image using low-level features and shape analysis. IEEE Trans. Inf. Forensics Secur. **11**(7), 1461–1475 (2016)
8. Csurka, G., Dance, C.R., Fan, L., Willamowski, J., Bray, C.: Visual categorization with bags of keypoints. In: Workshop ECCV, pp. 1–22 (2004)
9. ISO: Information technology biometric presentation attack detection part 3: Testing and reporting, JTC 1/SC 37, Geneva, Switzerland ISO/IEC FDIS 30107–3:2017 (2017)
10. Bosch, A., Zisserman, A., Munoz, X.: Image classification using random forests and ferns. In: IEEE International Conference on Computer Vision (2007)
11. Vedaldi, A., Fulkerson, B.: VLFeat: an open and portable library of computer vision algorithms (2008). http://www.vlfeat.org/
12. Vedaldi, A., Zisserman, A.: Efficient additive kernels via explicit feature maps. Pattern Anal. Mach. Intellingence **34**(3), 480–492 (2011)
13. Yambay, D., Ghiani, L., Denti, P., Marcialis, G.L., Roli, F., Schuckers, S.: Livdet 2011-fingerprint liveness detection competition 2011. In: 2012 5th IAPR International Conference on Biometrics (ICB), pp. 208–215. IEEE (2012)
14. Nogueira, R.F., de Alencar Lotufo, R., Machado, R.C.: Evaluating software-based fingerprint liveness detection using convolutional networks and local binary patterns. In: 2014 IEEE Workshop on BIOMS Proceedings, pp. 22–29. IEEE (2014)
15. Xia, Z., Lv, R., Zhu, Y., Ji, P., Sun, H., Shi, Y.-Q.: Fingerprint liveness detection using gradient-based texture features. Signal Image Video Process. **11**(2), 381–388 (2017)

16. Xia, Z., Yuan, C., Sun, X., Sun, D., Lv, R.: Combining wavelet transform and LBP related features for fingerprint liveness detection. IAENG Int. J. Comput. Sci. **43**(3), 290–298 (2016)
17. Jiang, Y., Liu, X.: Spoof fingerprint detection based on co-occurrence matrix. Int. J. SERSC **8**(8), 373–384 (2015)
18. Schuckers, S., Johnson, P.: Fingerprint pore analysis for liveness detection. US Patent App. 14/243,420, 2 April 2014
19. Jia, X., Yang, X., Cao, K., Zang, Y., Zhang, N., Dai, R., Zhu, X., Tian, J.: Multi-scale local binary pattern with filters for spoof fingerprint detection. Inf. Sci. **268**, 91–102 (2014)