Advanced Face Presentation Attack Detection on Light Field Database

Valeria Chiesa[1], Jean-Luc Dugelay[1]

Abstract: In the last years several works have been focused on the impact of new sensors on face recognition. A particular interest has been addressed to technologies able to detect the depth of the scene as light field cameras. Together with person identification algorithms, new anti-spoofing methods customized for specific devices have to be investigated. In this paper, a new algorithm for presentation attack detection on light field face database is proposed. While distance between subject and camera is not a relevant information for standard 2D spoofing attacks, it could be important when using 3D cameras. We prove through three experiments that the proposed method based on depth map elaboration outperforms the existent algorithms in presentation attack detection on light field images.

Keywords: Light field, Presentation attack detection, Anti-spoofing, Depth map.

1 Introduction

Automatic face recognition systems are nowadays used in different areas, from public security (border checkpoints, video-surveillance) to person identification for private purposes (unlock smartphones, customize smart cars) [Ra18, HV18]. The robustness against spoofing attacks is an crucial parameter for evaluating recognition system performances. In particular, the identification of attacks performed by presenting fake biometric traits to the sensor, namely presentation attacks detection (PAD), is deeply investigated in literature [RB17]. The development of new technologies pushes the need for updated methods customized for the specific sensors. In the last years, light field cameras rose the interest of the scientific community [GUC13, Se17, RRB15, JZW16, RB14]. Light field (or plenoptic) images are based on the idea that multiple information are collected for each pixel. Thus, in addition to light intensity, the camera is able to record the incoming direction of the light rays. From these data, a 3D representation of the scene could be rendered. Light field acquisitions can be obtained from different devices: we used a standard camera with a microlens array insert between the main lens and the sensor. A detailed dissertation about geometrical and physical properties of lens-less light field camera could be found in [LH96, Ng06]. Plenoptic raw data can be rendered in several ways, among which multi-view representation where several images are generated from a slightly shifted point of view and pair of RGB and depth map images.

The main contributions of this work are: a study of the light field imaging properties in relation to presentation attack detection; an introduction of a novel method able to exploit

[1] Department of Digital Security, EURECOM, Sophia-Antipolis, France, {chiesa, dugelay}@eurecom.fr

plenoptic characteristics in order to prevent spoofing attacks; the analysis of the performances of the proposed method in different training condition.

In sections 2 and 3, state of the art feature extractions and used database are described. In section 4 a new method is proposed. Three experiments and their results are shown in section 5. Finally, conclusions are presented.

2 State of the Art Features Specific to Light Field Data

Most of the works targeting presentation attack detection (PAD) analyze plenoptic data applying features extraction followed by a classification based on Support Vector Machine (SVM). PAD methods customized for light field images have already been proved to be more efficient on plenoptic data than algorithms created for standard RGB data [SMPC18]. For this reason, only features designed for light field data are compared with the proposed method. In particular, two feature vectors presented in [Se18] and [SMPC18] are chosen. Both methods are based on a multi-view representation of light field data. [Se18] describes a variation of the classical Local Binary Pattern (LBP) algorithm customized for light field images [AHP06] where, instead of considering adjacent pixels, values from different views transposed in the HVS and YC_bC_r color spaces are used. In this scheme, two classifiers are created, one trained with LBP_{HVS} features and one with $LBP_{YC_bC_r}$. Scores are merged in order to give the final classification result. The method shown in [SMPC18] is inspired to [Se18] but, instead of LBP algorithm, it emulates the well known Histogram of Oriented Gradients (HOG) method [DT05]. We have reproduced the code related to the first algorithm, while the second has been provided by authors.

3 Database

In order to test the performances of the proposed features, the database described in [Se18] is used. At the status of our knowledge, this is the only light field face database that includes raw data representing six different presentation attack modalities: printed paper, wrapped printed paper, laptop, tablet and two mobile models. The database consists of two sessions containing 50 subjects each one, for a total of 100 bona fide images and 600 presentation attack images. It is important to highlight that disparities among different views are dependent on the distance between the subject (or subject representation) and the camera and, typically, the closer is the subject, the higher is the disparity. In the face database presented in [Se18] such distance has not been annotated. In both [SMPC18] and [Se18] there is no mention about any data normalization useful to limit the impact of this factor.

4 Landmark Depth Features

While features used in [SMPC18] and [Se18] are based on the light field image property of being rendered as multi-views, our method exploits pair RGB-depth map representations.

4.1 Feature Computation

Raw data are elaborated with the proprietary software Lytro Desktop [Ly] and for each image a pair of RGB and depth picture is extracted. The software is able to set a perfect matching between RGB image and depth map so that a depth value is associated to each pixel of texture picture. First, landmark detection is performed with a method based on Histogram of Oriented Gradients (HOG). The implementation is described in [KS14] and the model is trained on the database described in [Sa16]. The algorithm identifies 68 landmarks for each human face in the image and detects a rectangular Region Of Interest (ROI). The size of the ROI depends on the dimension of the face represented in the image and in this case it does not impact in features computation. Then, for each landmark, the associated depth value is considered. In order to smooth eventual noise, the depth map is convoluted with a 7x7 pixels average filter. Landmark Depth Features (LDF) are defined as the set of depth values associated to the 68 landmarks Fig 1.

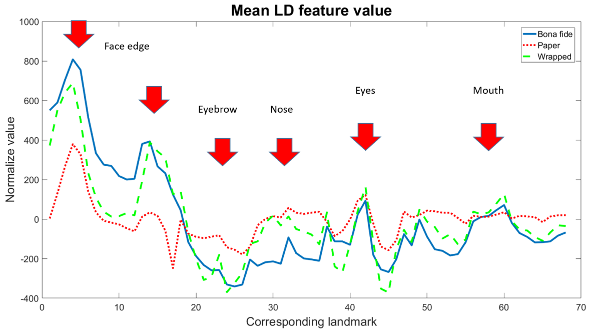

Fig. 1: Average LDF value. In order to be represented in the same graph, the mean value of each curve has been subtracted. On x axis is reported the number of correspondent landmark (landmarks order is not indicative. The corresponding map can be find in [KS14]).

The depth of particular landmarks could be more effective than others for detecting a presentation attack. Thus, with the purpose of investigating linear combination of landmark depth, a Principal Component Analysis (PCA) is performed. Fifty sets composed of an equal number of bona fide samples and of attack presentation samples (equally distributed among the possible attacks) are randomly created. Applying PCA to the whole database proves that more than 99.99% (evaluated as cumulative sum of eigenvalues) of information is stored in the first 10 principal components. Thus, Principal Landmark Depth Features (PLDF) are defined as the first 10 principal components computed as described.

4.2 A Preliminary Study

A preliminary study of depth maps shows an evident data clusterization according with attack type Fig 2. Previous analyses demonstrate that the light field depth information is deeply influenced by light conditions.

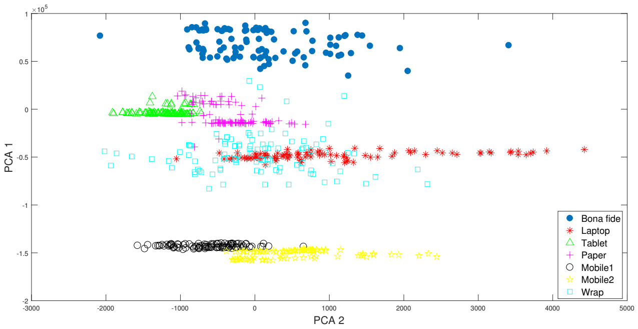

Fig. 2: Samples represented in 2D space created with the first two principal components. It is possible to observe a clear clusterization according with the first principal component.

Nevertheless, a focused investigation of screen backlight impact has not been carried out yet. Although we could not be sure about the reason of this clusterization, we believe that it may be ascribed to different distances between the subject and the camera during data acquisition. This information can be important in presentation attack detection. In fact, when the attack is performed with devices such as smartphone, the identification of spoofed images is straightforward: the size of the screen is considerate being smaller than a human face and the device must be held closer to the camera in order to present a credible image. Contrary to standard cameras, light field sensors can easily detect the proximity analyzing the depth map values. Thus, if face recognition system is based on plenoptic cameras, an eventual impostor faces an additional challenge: the size of spoofed face has to be plausible.

5 Experiments

Experiment performances are evaluated considering Bona fide Presentation Classification Error Rate (BPCER), Attack Presentation Classification Error Rate (APCER) and the average value of them called Average Classification Error Rate (ACER) [IS06]. During the training phase of all the experiments, a optimal threshold corresponding to Equal Error Rate (EER) point is defined. The reported values are evaluated on the test set using the threshold chosen in the training phase.

5.1 Experiment 1

In the first experiment, data are randomly divided in training (75% of samples) and test set (25% of samples). The samples used in training phase are not considered in the test set. The number of bona fide and artifact images is kept equal in the training set to avoid unbalanced results. A C-SVM with linear kernel implemented in [CL11] is used as classifier. One attack modality par time is processed, thus, six classifiers are created. The experiment is repeated 50 times with different compositions of training and test set in order to prevent overfitting. In Tab 1 the average percentage of error (ACER) is presented for each spoofing attack and for each considered feature set. ACER value is lower than 5% for all combinations of features and presentation attack modalities considered. LDF-based method is able to classify perfectly most of the attacks and only facing Wrapped Paper spoofing has a small percentage of failure ($< 0.75\%$).

Features	Pap	Tab	Lap	Mob1	Mob2	Wrap
$LBP_{HSV+YC_bC_r}$	0.68%	1.87%	0%	2.11%	0%	1.14%
HDG	0.80%	2.58%	0.16%	4.08%	2.27%	**0.28%**
LDF (Proposed method)	**0%**	**0%**	**0%**	**0%**	**0%**	0.46%
PLDF (Proposed method)	**0%**	**0%**	**0%**	**0%**	**0%**	0.72%

Tab. 1: Average ACER value evaluated over 50 runs of experiment 1.

5.2 Experiment 2

Often in a real world scenario, it is not possible to know in advance in which way the recognition system will be attacked. The second experiment is designed to create a classifier able to recognize all the attacks present in the database at the same time. As for the experiment described in 5.1, the training set contains an equal number of bona fide samples and artifact. Conversely, all the spoofing attack modalities in the database are considered together as artifact samples. The procedure is applied 50 times with different split of training and test set in order to avoid overfitting. The performances of the classifier (a C-SVM as described in 5.1) are reported for each attack. In Fig 3, DET curves for each presentation attack modality are shown. Average ACER values are presented in Tab 2. As for experiment 1, LDF method outperforms the algorithms proposed in [Se18] and [DT05].

Features	Pap	Lap	Tab	Mob1	Mob2	Wrap
$LBP_{HSV+YC_bC_r}$	2.09%	3.15%	3.98%	2.04%	4.17%	6.56%
HDG	3.96%	4.04%	19.08%	3.99%	3.96%	8.17%
LDF (Proposed method)	**0%**	**0%**	**0%**	**0%**	**0%**	**0.74%**
PLDF (Proposed method)	**0%**	**0%**	**0%**	**0%**	**0%**	0.8%

Tab. 2: Average ACER value evaluated over 50 runs of experiment 2.

5.3 Experiment 3

In the last experiment, no spoofed images are involved in training phase. A One-Class SVM trained only with bona fide images is used to identify all the presentation attacks included in the database. Starting from the elaboration of the information stored in only one class, in this case bona fide images, the One-Class classifier is able to discriminate test samples in two classes: belonging and not belonging to training class. As for experiments 5.1 and 5.2, the kernel of the SVM is linear; the parameter ν is chosen empirically for each set of features. Since the One-Class SVM implementation does not provide output scores, LBP_{HSV} and $LBP_{YC_bC_r}$ feature sets are tested separately. The size of the training set impacts

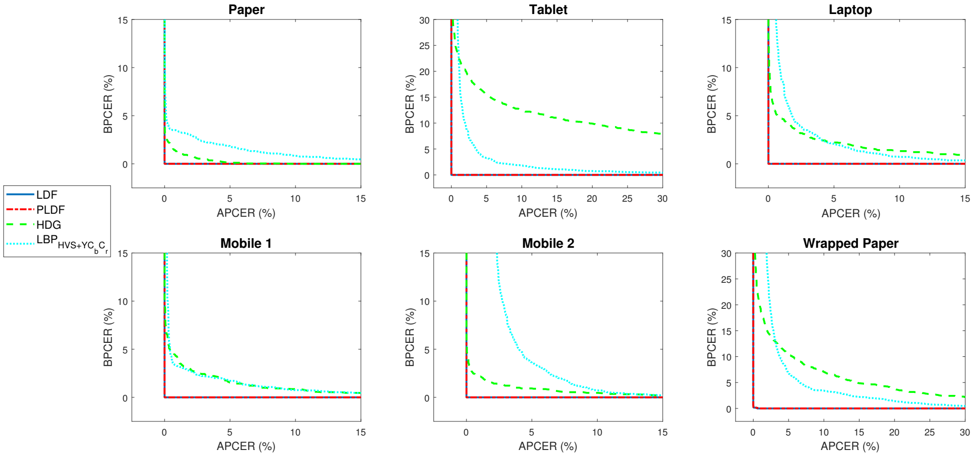

Fig. 3: DET curve for experiment 2. The proposed method (LDF) outperforms the other algorithms tested when evaluated on all considered presentation attacks. Also the performances of the presented reduct version (PLDF) reaches good results.

significantly on the stability of One-Class SVM. In Fig 4 is represented the percentage of ACER varying the training set size. While for the proposed method (LDF) a training set composed by 30 bona fide samples leads to high accuracy, HDG-based method reaches a stable accuracy with a bigger training set because of the higher number of features used (23400). The state of the art methods here represented do not perform as well as LDF and they may not be as versatile as the proposed one when changing the presentation attack. In particular LBP_{HSV} and $LBP_{YC_bC_r}$ performances decrease when the methods are seen separately [Se18].

6 Conclusion

In this work the problem of presentation attack detection in face recognition systems is tackled for plenoptic images. A new method based on light field images property of being rendered as pair of RGB and depth map is presented. A preliminary analysis shows how, contrary to 2D standard imaging, distance between subject and camera can impact on the performance of a system able to identify spoofed images. Thus, a presentation attack realized with a support smaller (or bigger) than an average face is easily detected: in fact, the impostor has to hold the fake representation at a different distance from the camera to show a credible face. Three experiments are carried out in order to study the characteristics of the proposed features in different training conditions. While in the first experiment, each presentation attack is separately analyzed, in the second all assaults are considered during the classifier training. The last experiment tests the possibility to use only bona fide sample to train a system able to detect any attack. All the experiments lead to almost perfect classification results obtained with the proposed method, outperforming the state

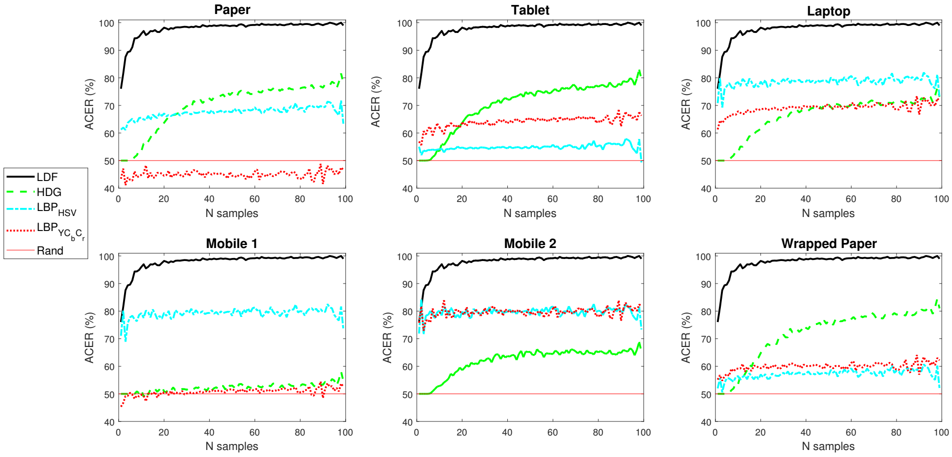

Fig. 4: Average ACER value varying training size. Convergence is reached with a small training set for all methods. In all studied cases, LDF algorithm obtains the higher ACER.

of the art approaches. This work shows how presentation attacks designed for standard 2D cameras are not effective in spoofing plenoptic sensors. Thus, it paves the way for the study of attacks customized for light field sensors. Moreover, a deeper investigation of distance subject-camera effect is becoming necessary to improve the analysis on face recognition on pleniptic images. With this aim, a new database acquired with a particular attention to distance subject-camera should be collected.

7 Acknowledge

This work has been partially supported by European Union's Horizon 2020 research and innovation program under grant agreement No 700259, (PROTECT).

References

[AHP06] Ahonen, T.; Hadid, A.; Pietikainen, M.: Face Description with Local Binary Patterns: Application to Face Recognition. IEEE Transactions on Pattern Analysis and Machine Intelligence, 28(12):2037–2041, Dec 2006.

[CL11] Chang, CC.; Lin, CJ.: LIBSVM: A library for support vector machines. ACM Transactions on Intelligent Systems and Technology, 2:27:1–27:27, 2011.

[DT05] 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). volume 1, pp. 886–893 vol. 1, June 2005.

[GUC13] GUCLF: a new light field face database, volume 8785, nov 2013.

[HV18] Hassan, A.; Viriri, S.: Invariant feature extraction for facial recognition: A survey of the state-of-the-art. In: 2018 Conference on Information Communications Technology and Society (ICTAS). pp. 1–6, March 2018.

[IS06] ISO: ISO/CEI 19795-1:2006: Information technology - Biometric performance testing and reporting - Part 1: Principles and framework. Iso, International Organization for Standardization, Geneva, Switzerland, 2006.

[JZW16] Ji, Z.; Zhu, H.; Wang, Q.: LFHOG: A discriminative descriptor for live face detection from light field image. In: 2016 IEEE International Conference on Image Processing (ICIP). pp. 1474–1478, Sept 2016.

[KS14] Kazemi, V.; Sullivan, J.: One millisecond face alignment with an ensemble of regression trees. In: 2014 IEEE Conference on Computer Vision and Pattern Recognition. pp. 1867–1874, June 2014.

[LH96] Levoy, M.; Hanrahan, P.: Light field rendering. SIGGRAPH '96 Proceedings of the 23rd annual conference on Computer graphics and interactive techniques, pp. 31–42, 1996.

[Ly] Lytro website. https://www.lytro.com/. Accessed: 2017-05-10.

[Ng06] Ng, R.: Digital Light Field Photography. PhD thesis, Stanford, CA, USA, 2006. AAI3219345.

[Ra18] Ranjan, R.; Sankaranarayanan, S.; Bansal, A.; Bodla, N.; Chen, J. C.; Patel, V. M.; Castillo, C. D.; Chellappa, R.: Deep Learning for Understanding Faces: Machines May Be Just as Good, or Better, than Humans. IEEE Signal Processing Magazine, 35(1):66–83, Jan 2018.

[RB14] Raghavendra, R.; Busch, C.: Presentation attack detection on visible spectrum iris recognition by exploring inherent characteristics of Light Field Camera. In: IEEE International Joint Conference on Biometrics. pp. 1–8, Sept 2014.

[RB17] Ramachandra, R.; Busch, C.: Presentation Attack Detection Methods for Face Recognition Systems: A Comprehensive Survey. ACM Comput. Surv., 50(1):8:1–8:37, March 2017.

[RRB15] Raghavendra, R.; Raja, K. B.; Busch, C.: Presentation Attack Detection for Face Recognition Using Light Field Camera. IEEE Transactions on Image Processing, 24(3):1060–1075, March 2015.

[Sa16] Sagonas, C.; Antonakos, E.; Tzimiropoulos, G.; Zafeiriou, S.; Pantic, M.: 300 Faces In-The-Wild Challenge: database and results. 47, 01 2016.

[Se17] Sepas-Moghaddam, A.; Chiesa, V.; Correia, P. Lobato; Pereira, F.; Dugelay, JL.: The IST-EURECOM light field face database. In: IWBF 2017, 5th International Workshop on Biometrics and Forensics, 4-5 April 2017, Coventry, UK. Coventry, UK, 04 2017.

[Se18] Sepas-Moghaddam, A.; Malhadas, L.; Correia, P. L.; Pereira, F.: Face spoofing detection using a light field imaging framework. IET Biometrics, 7(1):39–48, 2018.

[SMPC18] Sepas-Moghaddam, A.; Pereira, F. Vieira; Correia, P. Lobato: Light Field-Based Face Presentation Attack Detection: Reviewing, Benchmarking and One Step Further. IEEE Transactions on Information Forensics and Security, 13:1696–1709, 2018.