Received: 30 December 2021 | Revised: 4 June 2022 | Accepted: 5 August 2022

DOI: 10.1049/bme2.12092

IET Biometrics

The Institution of Engineering and Technology WILEY

ORIGINAL RESEARCH

Combining 2D texture and 3D geometry features for Reliable iris presentation attack detection using light field focal stack

Zhengquan Luo[1] | Yunlong Wang[2] | Nianfeng Liu[2] | Zilei Wang[1]

[1]Department of Automation, University of Science and Technology of China, Hefei, Anhui, China

[2]Center for Research on Intelligent Perception and Computing (CRIPAC), National Laboratory of Pattern Recognition (NLPR), Institute of Automation, Chinese Academy of Sciences (CASIA), Beijing, China

Correspondence

Yunlong Wang, Center for Research on Intelligent Perception and Computing (CRIPAC), National Laboratory of Pattern Recognition (NLPR), Institute of Automation, Chinese Academy of Sciences (CASIA), No.95, Zhongguancun East Street, Haidian District, Beijing 100190, China.
Email: yunlong.wang@cripac.ia.ac.cn

Funding information

CAAI Huawei MindSpore Open Fund, Grant/Award Number: CAAIXSJLJJ-2021-053A; National Natural Science Foundation of China, Grant/Award Numbers: 61906199, 62006225, 62176025; Strategic Priority Research Program of Chinese Academy of Sciences, Grant/Award Number: XDA27040700

Abstract

Iris presentation attack detection (PAD) is still an unsolved problem mainly due to the various spoof attack strategies and poor generalisation on unseen attackers. In this paper, the merits of both light field (LF) imaging and deep learning (DL) are leveraged to combine 2D texture and 3D geometry features for iris liveness detection. By exploring off-the-shelf deep features of planar-oriented and sequence-oriented deep neural networks (DNNs) on the rendered focal stack, the proposed framework excavates the differences in 3D geometric structure and 2D spatial texture between bona fide and spoofing irises captured by LF cameras. A group of pre-trained DL models are adopted as feature extractor and the parameters of SVM classifiers are optimised on a limited number of samples. Moreover, two-branch feature fusion further strengthens the framework's robustness and reliability against severe motion blur, noise, and other degradation factors. The results of comparative experiments indicate that variants of the proposed framework significantly surpass the PAD methods that take 2D planar images or LF focal stack as input, even recent state-of-the-art (SOTA) methods fined-tuned on the adopted database. Presentation attacks, including printed papers, printed photos, and electronic displays, can be accurately detected without fine-tuning a bulky CNN. In addition, ablation studies validate the effectiveness of fusing geometric structure and spatial texture features. The results of multi-class attack detection experiments also verify the good generalisation ability of the proposed framework on unseen presentation attacks.

1 | INTRODUCTION

Iris liveness detection is an indispensable module of any iris recognition system that blocks spoof attacks from malicious entities. With the continuous development of the iris PAD arms race, diverse presentation attack instruments (PAIs) are constantly evolving, which puts more pressure on guaranteeing the security of iris recognition systems.

In the literature, software-based iris liveness detection methods usually rely on subtle differences in colour, texture, and context to distinguish bona fide and spoofing iris samples. The spectral domain [1] and iris quality criteria [2] are first investigated. Various local-based descriptors are employed to detect spoofing iris samples such as LBP [3], BSIF [4, 5], and GLCM [6]. Higher detection accuracy was achieved by fusing multiple local features [7, 8]. In addition, the spatial pyramid and relational measures are employed by RegionalPAD [7] to extract the features from local neighbourhoods. Other recent SOTA methods can be reviewed in the recent competition of iris liveness detection, that is, LivDet-Iris 2020 [9]. In this competition, some convolutional neural network (CNN)-based methods are introduced into iris liveness detection. For instance, MTCNN [10] is employed by Chen and Ross [11]. They modified it and proposed MTPAD for an automatic iris presentation attack detection solution. D-NetPAD [12] adopts DenseNet161 [13] as the backbone architecture, which outperforms other methods in this competition, and its performance was not satisfactory on proprietary test data. In Ref. [14], a cascade of MobileNetV2 [15] modifications were trained from scratch and utilised to recognise PAIs, which achieved

promising performance on the LivDet-Iris 2020 competition database. The bottleneck of these methods is that the missing stereoscopic characteristics of human eyes in 2D planar images make it intractable to consider the discrepancy in 3D geometric structure between bona fide and presentation attack iris samples. On the other hand, extra equipment could be utilised in hardware-based iris liveness detection methods to analyse the inherent properties of the living iris. For example, Daugman [1] investigated Purkinje reflection of the cornea and lens in the eye. Lee et al. [16] solved the liveness detection task through reflectance changes between the sclera and iris under different near-infrared light. Czajka [17] detected the living iris by the dynamic changes of the pupil for different illuminations. However, these approaches need to add an extra device, which usually costs much and requires capturing iris dynamics. Hence, they are often infeasible in practical use.

On the other hand, with the aid of an internal microlens array (MLA), LF cameras are able to capture 4D spatial-angular information in a single photographic exposure, which contains huge information on the 3D geometric structure and reflectance property of the object surface. Actually, the differences in the 3D surface geometry and reflectance between bona fide and presentation attack iris samples are the intrinsic properties that PAD methods can exploit. Flat iris PAIs including printed papers, printed photos, and electronic displays exhibit limited distinctions in the depth layout of ocular regions. PAIs that simulate 3D geometric structure, such as prosthetic eye balls, are thus more challenging when considering only depth/disparity analysis. However, these 3D-simulation PAIs inevitably transform the iris texture as a consequence of changes in the shadows and contrast gradients. In other words, light field imaging gains huge advantages over other hardware-based solutions in exploring intrinsic features for iris PAD. Raghavendra and Busch [18] first introduced LF imaging into iris liveness detection. They captured real and artificial iris images under the visible wavelength (VW) spectrum from 104 unique samples using a commercial *Lytro* camera and exploited the variation of focus between multiple digitally refocussing images. Some approaches have been proposed to adopt LF imaging for face anti-spoofing by designing specific handcrafted features [19–24], and more related work can be reviewed in Ref. [25]. Recently, Liu et al. [26] began to utilise a CNN in LF-based face liveness detection. Song et al. [27] built a middle-sized dataset containing 504 bona fide and spoofing near-infrared (NIR) iris samples using a lab-made microlens-based LF camera. The focal stack is rendered from LF images by refocussing at a group of depth layers. They combined the focus energy value and LPQ feature extracted from LF focal stack to defend iris spoof attacks. Obviously, the number of LF images in the public domain is far from sufficient to fine-tune or retrain data-hungry deep learning frameworks for iris liveness detection.

In this paper, we investigate intrinsic features of iris PAD by exploring 3D geometry and 2D texture differences in the LF focal stack. Differences in depth layout and texture are internally reflected in the defocus blur and local patterns between different rendered slices of the focal stack. Additionally, we attempt to handle the PAD dilemma without collecting more data on spot, which focusses on a general off-the-shelf solution for iris liveness detection. The proposed framework adopts a sequence-oriented model to extract 3D geometric structure and a planar-oriented model to extract 2D spatial texture feature from LF focal stack, respectively. The adopted sequence-oriented model includes C3D [28] model pre-trained on Sports 1M [29] and P3D [30] model pre-trained on Kinetics 400 [31]. The input of the sequence-oriented model is the sequence of focal stack slices refocussing at different depth layers. The adopted planar-oriented model includes ResNet50 [32], InceptionV3 [33], and MobileNetV2 [15] pre-trained on ImageNet [34]. The input of planar-oriented model is the sharpest slice selected from the focal stack via focus level assessment. Although these two models are specifically trained for other vision tasks and have not seen any LF iris data, it is experimentally verified that the off-the-shelf CNN features are discriminative for iris liveness detection. Classified by a support vector machine (SVM) classifier, both the off-the-shelf deep features yielded from planar and sequence oriented pre-trained models can achieve better performance than the handcrafted methods with the same input. Various presentation attacks, including printed papers, printed photos, and electronic displays, can be accurately detected. Furthermore, these features are fused in a two-branch manner. Ablation studies prove that 3D geometric structure and 2D spatial texture features are complementary in iris liveness detection. By fusing two kinds of features, the proposed framework can not only obtain higher accuracy but also gain superiority in resisting the degradation caused by motion blur, noise, and other low-quality factors. In summary, the proposed method is of considerable reliability and robustness, which can be directly transplanted to iris recognition systems and defend against various presentation attacks.

2 | FRAMEWORK

An overview of the proposed framework is depicted in Figure 2, and the details will be elaborated in this section.

2.1 | Light field focal stack

4D light field data can be decoded from LF images and expressed as the two-plane parameterisation model $L(u, v, s, t)$. The traditional 2D planar image can be obtained by integrating the LF function as $I(s, t)$:

$$I(s,t) = \iint L(u,v,s,t) \cdot du \cdot dv \tag{1}$$

LF cameras can simultaneously capture the light intensity and directional information in a single exposure. Therefore, the focus plane can be changed at any depth layers by the digital refocussing method proposed by Ng et al. [35]. L' represents the synthetic film plane, and L is the original microlens plane. α

indicates the relative locations of these two planes: $\alpha = F'/F$. The schematic of digital refocussing is depicted in Figure 1 and can be derived as (2):

$$L'(u,v,s',t') = L\left(u,v,\frac{s'}{\alpha}+u(1-\alpha),\frac{t'}{\alpha}+v(1-\alpha)\right) \quad (2)$$

2.2 | 2D spatial texture feature

The 2D spatial texture feature extracted from the best focus slice of the iris region in the focal stack is more appropriate than other slices for liveness detection because the differences in texture details between bona fide and presentation attack iris images are more significant. The Tenengrad gradient variance (TGV) [36] is employed here as the focus measure function F_{TGV}.

The computation of the TGV value of each slice I_α in the focal stack is given as in (3):

$$\bar{I}_\alpha(s,t) = \sqrt{(I_\alpha(s,t) * E_s)^2 + (I_\alpha(s,t) * E_t)^2}$$

$$m = \frac{1}{ST}\sum_{s=1}^{S}\sum_{t=1}^{T}\bar{I}_\alpha(s,t) \quad (3)$$

$$F_{TGV}(I_\alpha) = \sum_{s=1}^{S}\sum_{t=1}^{T}\left[I^*_\alpha(s,t) - m\right]^2$$

The spatial resolution of each slice in the focal stack is $S \times T$. The Sobel operators E_s and E_t are used to extract edge information. The focus energy of the sharpest slice of the iris region is the maximum in the focal stack, and the focus energy curve usually presents a single peak. The best focus plane of the iris region is thus obtained via (4), and the sharpest iris image is expressed as I_{α^*}.

$$\alpha^* = \underset{\alpha\in[\delta_1,\delta_2]}{argmax}\{F_{TGV}(I_\alpha)\} \quad (4)$$

A simple coarse-to-fine strategy is applied to search the best focus plane. Specifically, $[\delta_1, \delta_2]$ in Equation (4) is first set to a large range and then a coarse selection $\alpha^\dagger$ is derived. Next, the search range is narrowed down to $[\alpha^\dagger - \varepsilon, \alpha^\dagger + \varepsilon]$ where ε is a small value. In this manner, the best focus location α^* can be determined more precisely. A planar-oriented model is utilised to extract 2D spatial texture feature from the sharpest iris image I_{α^*} as given in (5):

$$f_{2D} = F_{2D}(I_{\alpha^*}) \quad (5)$$

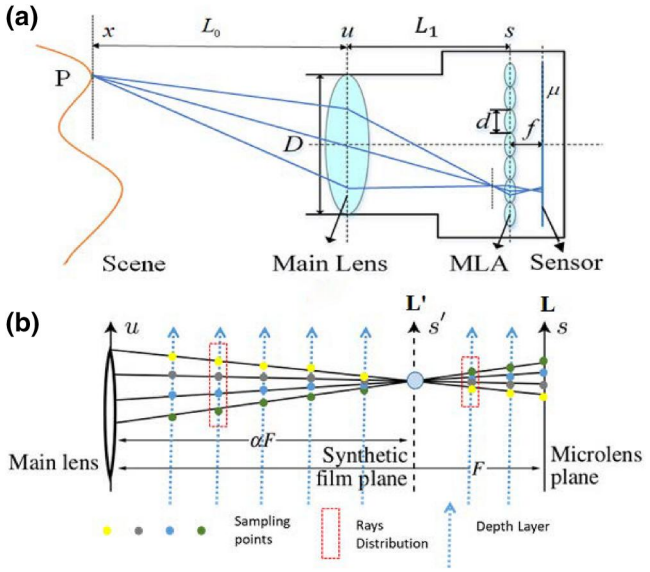

FIGURE 1 The schematic of digital refocussing

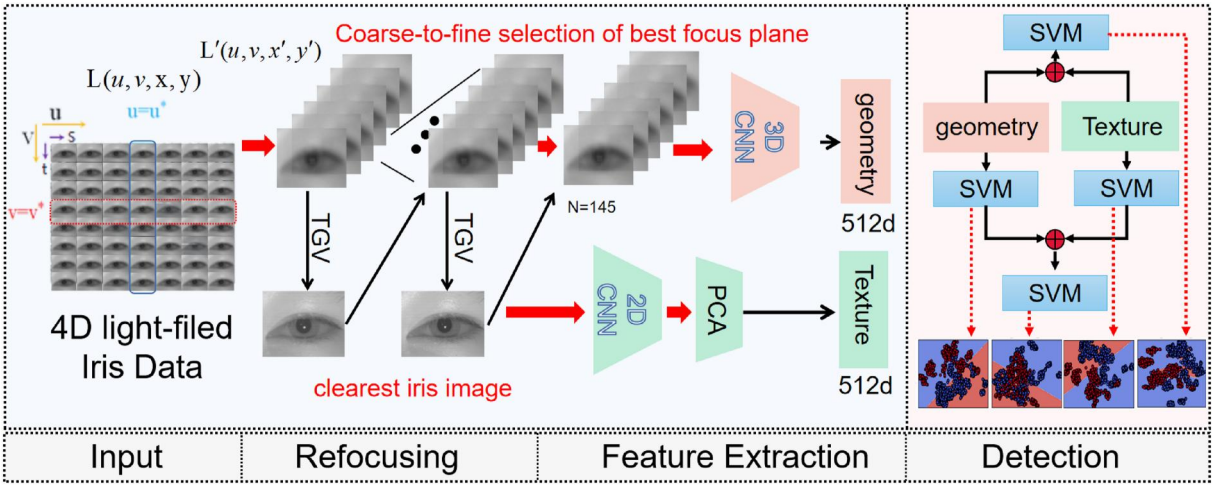

FIGURE 2 Overview of the proposed framework

2.3 | 3D geometric structure feature

After α^* is determined, a LF focal stack $\{I_\alpha\}$ that is uniformly distributed within a certain depth range in the vicinity of α^* is obtained by digital refocussing. The best focus slice of the iris region in the focal stack is set as the centre, which is presented as $\alpha \in [\alpha^* - 0.2, \alpha^* + 0.2]$. The step is $\Delta\alpha = 0.0028$, and the number of slices in the rendered focal stack is $N_{fs} = 145$.

The focal stack reveals the 3D geometric structure of the imaged objects. Different parts of the iris region will exhibit different focussing levels as α varies. Even though the forged 2D texture details of presentation attack irises can be similar to genuine iris samples, it is nearly impossible to fabricate the 3D geometric structure identical to real iris physiology. On the other hand, the entities of current presentation attacks such as printed papers, printed photos and electronic screen are usually flat or curved, which exhibit an almost constant amount of defocus blur between adjacent slices unlike bona fide images. The distinctions can be reflected in the difference between two adjacent slices in the focal stack as given in Equation (6):

$$\begin{aligned} I_{\Delta\alpha}(s,t) &= I_{\alpha+\Delta\alpha}(s,t) - I_\alpha(s,t) \\ &\approx \iint -\Delta\alpha \left[\left(\frac{s}{\alpha^2} + u\right)\frac{\partial}{\partial s} + \left(\frac{t}{\alpha^2} + v\right)\frac{\partial}{\partial t} \right] \\ &\quad \times L\left(u, v, \frac{s}{\alpha} + u(1-\alpha), \frac{t}{\alpha} + v(1-\alpha)\right) \Big] du dv \end{aligned} \tag{6}$$

Currently, the accuracy of pixel-wise depth or disparity estimation from an LF focal stack is limited. Instead, a sequence-oriented model is used to extract the 3D geometric structure feature of the iris region from the focal stack as given in (7). The pre-trained model is employed here to implicitly model the 3D geometric information in the focal stack, slightly analogous to exploiting spatiotemporal features from video clips for action recognition in Ref. [28].

$$f_{3D} = F_{3D}(\{I_\alpha\}) \tag{7}$$

2.4 | Two-branch feature fusion

Texture and geometric structure are two dominant factors of the difference between bona fide and spoofing irises. Nearly all presentation attacks have flaws in either or both of these two aspects when forging genuine iris. For instance, a spoofing iris displayed on an iPad screen may exhibit high resolution and clear texture details, but the geometric properties and reflections of human eyes are very different. The geometric structure of the prosthetic iris is close to that of the genuine iris, but the forged texture details and sharpness levels are dissimilar. Thus, it makes sense to fuse both 2D spatial texture and 3D geometric structure features for better iris liveness detection. The fusion of these two complementary features can enable the proposed framework to reliably defend more types of presentation attacks. Furthermore, it is experimentally verified that the robustness of the framework can be enhanced on resisting the degradation factors during data collection.

The 3D geometric structure feature vector f_{3D} and 2D spatial texture feature vector f_{2D} are first extracted by the respective branch, respectively. The dimension of f_{3D} extracted from C3D [28] and P3D [30] is 512. Meanwhile, the dimension of f_{2D} extracted from ResNet50 [32] and InceptionV3 [33] is 2048 while that of MobileNetV2 [15] is 1280. For consistency, the dimension of f_{2D} is reduced from its original length to 512 as f'_{2D} through principal component analysis (PCA), the length of which is thus the same as f_{3D}. Otherwise, the inconsistency of the dimensions will cause an imbalance in the contributions of each feature. After that, f_{3D} and f'_{2D} are concatenated together, and an SVM classifier is applied to classify the fused features into bona fide or presentation attack irises. In addition, the performance of f_{3D} only, f'_{2D} only, feature-level and score-level fusion are also validated as shown in the right side of Figure 2. The results of these ablation experiments are detailed in Section 3.3.

3 | DATASET AND EXPERIMENTS

3.1 | Dataset

The experiments are conducted on the dataset collected by Song et al. [27]. The dataset was captured using a lab-produced microlens-based LF camera and a commercial device IKUSB-E30 (http://www.irisking.com/pron.php?id=523) under NIR illumination. The setup of dataset collection was shown in Figure 3. The types of presentation attacks include printed papers, printed glossy photos, and electronic displays. The high-quality iris samples were first captured by IKUSB-E30, and then these high-quality samples were printed on papers and photos or displayed on the screen of iPad mini 4 to generate the artefacts. The main lens of the lab-produced LF camera was tuned to be in focus at a position of 1.6 m. Simultaneously, both bona fide and presentation attack iris samples were captured when the subjects and PAIs were standing at or be placed at three distances, that is, 1.5, 1.6, and 1.7 m. The dataset contains 504 samples from 14 subjects, consisting of 230 LF images of bona fide iris and 274 LF images of spoofing iris. The respective sample number of the PAIs, that is, printed papers, printed photos, and electronic display are 18, 122, and 134. The serious data imbalance among these PAIs intensifies the difficulty of PAD task. An example of raw LF image containing both eyes printed on photos is shown in Figure 4. Hexagonal microlens images can be observed from the close-up of iris in the raw LF image. The toolbox released by Dansereau et al. [37] was utilised to decode raw LF images into 4D LF data. The eye regions were cropped from the same location of each SAI. The spatial resolution of each SAI after cropping is 128 × 96, and the angular resolution is 7 × 7. Examples from the same subject's right eye in the dataset are shown in Figure 5. As described in Section 2.3, the

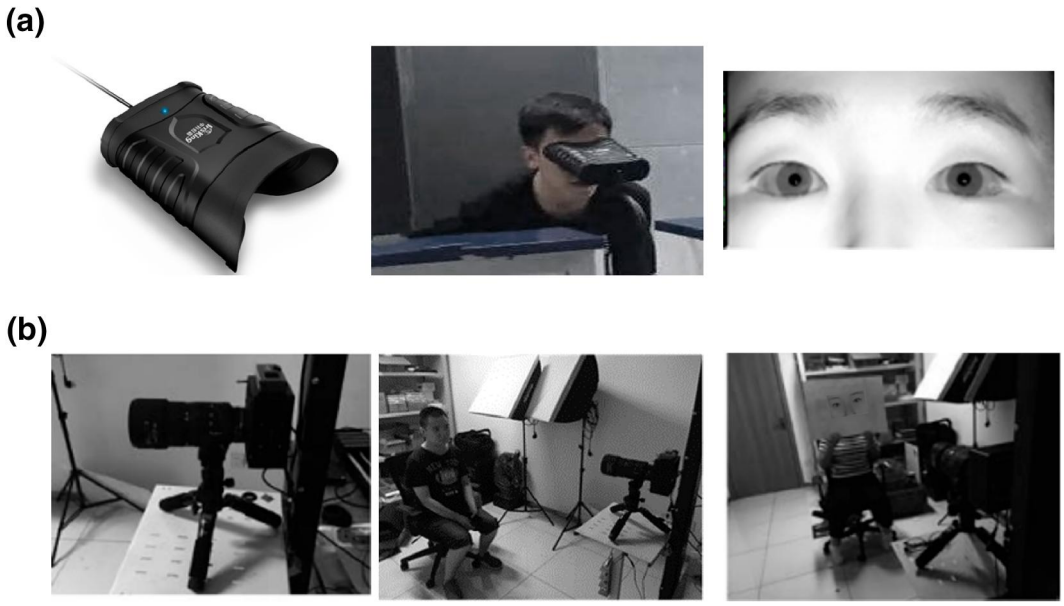

FIGURE 3 Setup of dataset collection. (a) The commercial device IKUSB-E30, in-person image capture and exemplar high-quality iris image. (b) The lab-produced light field (LF) camera, image capture of bona fide samples, and image capture of spoofing samples printed on glossy photos

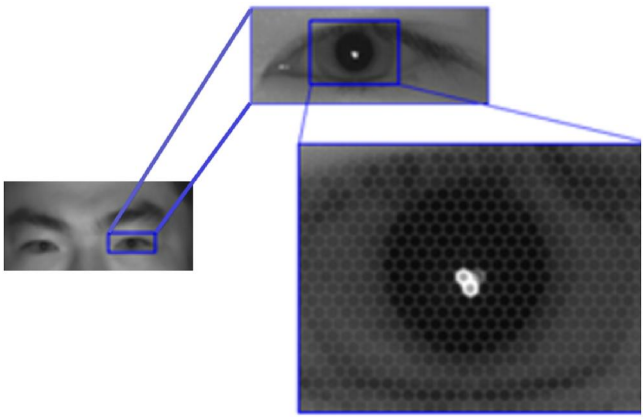

FIGURE 4 An example of raw light field (LF) image containing both eyes printed on photos. Hexagonal microlens images can be observed from the close-up of iris in the raw LF image

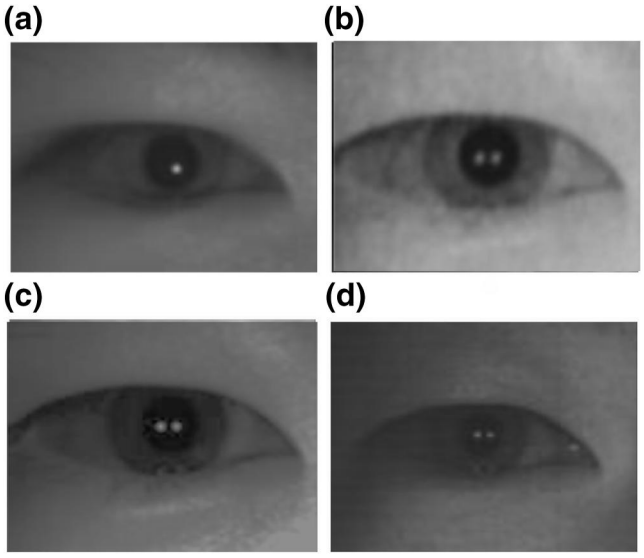

FIGURE 5 Examples from the same subject's right eye in the dataset. (a) Bona fide iris sample. (b) A4 paper printed iris sample. (c) Glossy photo printed iris sample. (d) Electronically displayed iris sample

rendered focal stack via digital refocussing has 145 slices around the best focus plane. The details of the adopted database is listed in Table 1. We have got the authority from the authors of Song et al. [27] and released the LF focal stack data on our website (http://www.cripacsir.cn/dataset/).

3.2 | Evaluation metrics

According to ISO/IEC IS 30107-3 [38], the evaluation metrics of iris liveness detection are:

I. Attack presentation classification error rate (APCER). APCER is the rate at which spoofing iris samples are mistakenly identified as bona fide samples.
II. Bona fide presentation classification error rate (BPCER). BPCER is the rate at which bona fide iris samples are identified as spoofing attacks erroneously.
III. Average classification error rate (ACER). ACER is the arithmetic mean of APCER and BPCER. The smaller the ACER, the better the performance of the iris liveness detection method. Although ACER has been deprecated in the latest publication of ISO/IEC 30107-3, it is still computed for the purpose of comparing with other former iris PAD methods.

TABLE 1 Details of the adopted database

Setup of image collection	
High-quality iris samples	IKUSB-E30
Bona fide and spoofing samples	Lab-produced LF camera
Focus location of LF camera	1.6 m
Distances of subjects or PAIs	1.5, 1.6, and 1.7 m
Illumination	NIR
Statistics of the database	
Spatial resolution of SAI	128 × 96
Angular resolution of LF sample	7 × 7
Length of LF focal stack	145
Number of subjects	14
Kinds of PAIs	3
Number of bona fide samples	230
Number of spoofing samples	274
Total number of images	73,080
Image distribution	
Bona fide	230
Printed papers	18
Printed photos	122
Electronic display	134

Abbreviations: NIR, near-infrared; PAIs, presentation attack instruments.

IV. $BPCER_{10}$ and $BPCER_{20}$. $BPCER_{10}$ is the BPCER when the APCER is fixed at 10% and $BPCER_{20}$ is the BPCER when the APCER is fixed at 5%.
V. Equal Error Rate (EER) is the operational point when the APCER is equal to the BPCER in the Detection Error Trade-off (DET) curve.

3.3 | Experiments

To verify the efficacy of the proposed method, four groups of experiments are conducted: 1) The proposed framework is compared with several iris liveness detection methods that operate on a 2D planar image or LF focal stack. 2) In ablation study, SVM classifiers with different kernels are tested on different fusion strategies, including 2D texture features only, 3D geometry features only, feature-level fusion and score-level fusion. The purpose is to verify the effectiveness of feature fusion. 3) A multi-class classification experiment verifies that the proposed framework can distinguish different presentation attacks. 4) The last experiment verifies the robustness of the fused features against degradation factors, such as motion blur and random noise.

K-fold cross-validation is adopted to optimise the SVM classifiers, where $K = 5$. In each cross-validation, 20% of the data with a mix of all artefact types available in the database are used to tune the parameters of SVM classifiers, and the other 80% of the data are used to validate the performance. Both the mean and standard deviation of evaluation metrics are reported according to the results of 5-fold cross validation. The SVM package of Scikit-learn [39] and MindSpore platform (https://www.mindspore.cn/) were adopted in the implementations. Some hyperparameters of the SVM classifier are selected through grid search and then set as $tol = 1e - 6$ and $C = 1000.0$. In addition, the kernels of the SVM classifier are selected as RBF. To reproduce the results in this paper, the source codes have been released on Github (https://github.com/luozhengquan/LFLD).

Comparative Experiment Table 2 tabulates the comparisons between variants of the proposed framework and several recent iris liveness detection methods that operate on 2D planar images or LF focal stacks, along with the results of the single texture or geometry branch. The DET curves of the proposed framework and other iris PAD methods are also plotted in Figure 6. In Table 2, APCER and BPCER were determined at a threshold of 0.5. $BPCER_{10}$, $BPCER_{20}$ and EER are independent of decision thresholds. Note that the results of $BPCER_{10}$, $BPCER_{20}$ and EER of some compared methods are NaN. It is due to that the source codes of these methods are not available and we directly adopt the nominal results from Ref. [27]. Besides, the mean and standard deviation of the evaluations are listed upon 5-fold cross validation. The variants of the proposed framework utilised different pre-trained DL models for off-the-shelf deep feature extraction, and only the parameters of SVM classifiers were optimised under the same dataset splitting. They are donated as *ResNet50 + C3D*, *ResNet50 + P3D*, *InceptionV3 + C3D*, *InceptionV3 + P3D*, *MobileNetV2 + C3D* and *MobileNetV2 + P3D*. Specifically, *ResNet50* [32], *InceptionV3* [33] and *MobileNetV2* [15] are pre-trained on ImageNet [34]. *C3D* [28] is pre-trained on Sports 1M [29] while *P3D* [30] is pre-trained on Kinetics 400 [31]. Note that *ResNet50*, *InceptionV3* and *MobileNetV2* also denote 2D texture branch while *C3D* and *P3D* denote 3D geometry branch. For fairness, DIIVINE [40] and LPQ [41] take the clearest iris image in the focal stack as input. DIIVINE [40] is based on image quality evaluation, and its ACER is 11.43%. LPQ is a local phase quantisation descriptor for texture analysis, and its ACER is 9.63%. The performance of Ref. [18] declines sharply on the NIR dataset because it only employs one empirically determined threshold to distinguish the bona fide and various PAIs. Ref[27] utilised manually crafted feature descriptors of the focus energy curve for LF-based iris liveness detection, and its ACER is 3.69%. In addition, we compared two recent SOTA methods, that is, MTPAD [11] and D-NetPAD [12], which achieved superior performances on the LivDet-Iris 2020 [9] database. The performance of MTPAD [11] and D-NetPAD [12] without fine-tuning on the dataset is very poor because the model trained on other PAD datasets is highly correlated with the corresponding data distribution. To further prove this statement, D-NetPAD [12] was fine-tuned on the same portion of the adopted database and tested on the rest. The fine-tuned version of the D-NetPAD [12] model is named as *FT_D-*

TABLE 2 Comparisons between variants of the proposed framework and several recent iris presentation attack detection (PAD) methods that operate on 2D planar images or light field (LF) focal stacks, along with the results of single texture or geometry branch.

	APCER (%)	BPCER (%)	ACER (%)	$BPCER_{10}$ (%)	$BPCER_{20}$ (%)	EER (%)
DIIVINE [40]	5.95	16.91	11.43	/	/	/
LPQ [41]	11.90	7.35	9.63	/	/	/
Raghavendra *et al.* [18]	32.14	50.74	41.44	/	/	/
Song *et al.* [27]	2.98	4.41	3.69	/	/	/
MTPAD [11]	40.01	39.78	39.90	/	/	/
D-NetPAD [12]	17.89 ± 4.75	66.96 ± 6.52	45.43 ± 5.64	73.48 ± 10.87	78.26 ± 15.22	45.64 ± 4.99
FT_D-NetPAD [12]	3.62 ± 4.52	5.78 ± 4.57	4.70 ± 4.55	0.91 ± 0.45	3.96 ± 2.67	4.17 ± 2.26
ResNet50 [32]	1.48 ± 1.18	0.96 ± 0.38	1.22 ± 0.78	0.06 ± 0.50	0.19 ± 0.27	1.24 ± 0.50
InceptionV3 [33]	2.21 ± 1.02	3.88 ± 3.00	3.05 ± 2.01	0.93 ± 1.04	2.18 ± 2.05	2.90 ± 1.50
MobileNetV2 [15]	2.19 ± 1.82	1.52 ± 0.87	1.85 ± 1.34	0.02 ± 0.02	0.33 ± 0.43	1.81 ± 1.05
P3D [30]	4.71 ± 0.68	1.78 ± 0.60	3.25 ± 0.64	0.29 ± 0.38	1.54 ± 0.82	3.29 ± 0.60
C3D [28]	4.07 ± 1.87	3.72 ± 3.09	3.90 ± 2.48	0.52 ± 0.65	3.82 ± 2.66	4.25 ± 1.22
ResNet50 + P3D	4.62 ± 0.73	1.78 ± 0.65	3.20 ± 0.69	0.26 ± 0.38	1.52 ± 0.81	3.23 ± 0.82
InceptionV3+P3D	4.65 ± 0.68	1.78 ± 0.64	3.22 ± 0.66	0.28 ± 0.38	1.50 ± 0.38	3.25 ± 0.64
MobileNetV2+P3D	4.65 ± 0.73	1.76 ± 0.65	3.21 ± 0.69	0.26 ± 0.38	1.61 ± 0.98	3.27 ± 0.59
ResNet50 + C3D	1.44 ± 1.18	**0.76 ± 0.70**	**1.10 ± 0.94**	0.04 ± 0.10	0.13 ± 0.27	1.15 ± 0.46
InceptionV3+C3D	2.11 ± 1.00	3.70 ± 2.99	3.01 ± 1.99	0.43 ± 0.33	2.17 ± 2.06	2.88 ± 1.49
MobileNetV2+C3D	**0.99 ± 1.28**	1.39 ± 1.79	1.19 ± 1.54	**0.02 ± 0.02**	**0.03 ± 0.03**	**1.05 ± 0.73**

Note: Best results are in bold values.

Abbreviations: ACER, Average classification error rate; APCER, Attack presentation classification error rate; BPCER, Bona fide presentation classification error rate; EER, Equal Error Rate.

NetPAD. Not surprisingly, its mean EER decreases from 45.64% to 4.17%. Dependence on training data greatly hinders the generalisation ability of DL-based iris PAD methods.

Compared with these methods, the best-performing variant of the proposed framework, that is, *MobileNetV2 + C3D* achieves a mean ACER of 1.19% and a mean EER of 1.05%. The proposed method outperforms other approaches by a large margin and even obviously surpasses *FT_D-NetPAD* in Table 2. It is also demonstrated by the DET curve depicted in Figure 6(a). It should be emphasised that the proposed method only utilises a small portion of the dataset (i.e. 20%) to adjust several hyperparameters of the SVM classifier. The process of optimising the SVM classifier is much more time-efficient and resource-friendly than training millions of weight parameters in *FT_D-NetPAD*.

Ablation study It can also be seen from Table 2 that the overall performance of two-branch fusion is obviously better than either of the single branches. Take *MobileNetV2 + C3D* for instance, the mean EER of *MobileNetV2* and *C3D* are 1.81% and 4.25%, respectively. Apparently, *C3D* is inferior in extracting discriminative representations for detecting iris PAIs. But fusion of these off-the-shelf deep features along with SVM classifier optimisation, that is, *MobileNetV2 + C3D*, obtains a mean EER of 1.05%. Undoubtedly, it indicates that the complementarity of 2D texture and 3D geometric features can be effectively explored by feature-level fusion. Laterally, the variants that combine *C3D* with 2D texture features significantly outperform those that combine *P3D* with 2D texture features.

In addition, the score distributions of the bona fide and spoofing iris samples output by *ResNet50 + C3D* are depicted in Figure 7. In Figure 7, texture feature means the output of *ResNet50* and geometry feature means the output of *C3D*. Feature fusion means the output of *ResNet50 + C3D*, which combines the two branches at feature level. Score fusion means the average of the outputs of *ResNet50* and *C3D*, which fuses the two branches at score level. It reveals that both feature-level and score-level fusion can provide more separate boundaries for classification. These results verify that both feature-level or score-level fusion of two branches can improve the performance of detecting PAIs.

Robustness analysis Figure 8 shows the curve of performance fluctuation of *MobileNetV2 + C3D* as a result of adding diagonal motion blur or zero-mean Gaussian noise. In the robustness analysis, the maximum degree of added motion blur D_{blur} is chosen from [50, 100, 150] in order. Then each slice in the focal stack is motion blurred diagonally with a randomly selected degree in the range of (0, D_{blur}]. Similarly, the maximum variance of added zero-mean Gaussian noise $D_{Gaussian}$ is chosen from [0.025, 0.050, 0.075]. Then, the variance of Gaussian noise is determined by randomly selecting from (0, $D_{Gaussian}$]. Finally the zero-mean and

variance-determined Gaussian noise is added to each slice in the focal stack. The performance fluctuation of the proposed framework against motion blur and Gaussian noise on the adopted dataset under 5-fold cross validation is depicted in Figure 8.

With the addition of motion blur or Gaussian noise, the performance drop of *MobileNetV2 + C3D* is rather smooth. As the level of added blur or noise level increases, the fluctuation of the mean EER of *MobileNetV2 + C3D* stays stable. The results demonstrate the robustness of the proposed framework against the degradation factors, which is usually a necessity of real-world applications.

Multi-class attack analysis A major concern in deploying an iris PAD solution is that there is no way to know what type of attacks will be performed beforehand. To further validate the proposed framework, the multi-class attack classification experiment applies the proposed framework to classify various presentation attacks. The leave-one-out protocol is employed to verify the efficacy of the proposed framework on unseen attacks. Specifically, all the samples belonging to one kind of PAIs and half of the bona fide iris samples are retained for testing, while all the other spoofing samples and the remaining bona fide samples are used for training. The best-performing variant *MobileNetV2 + C3D* is validated. The mean evaluation metrics of the PAIs are also reported.

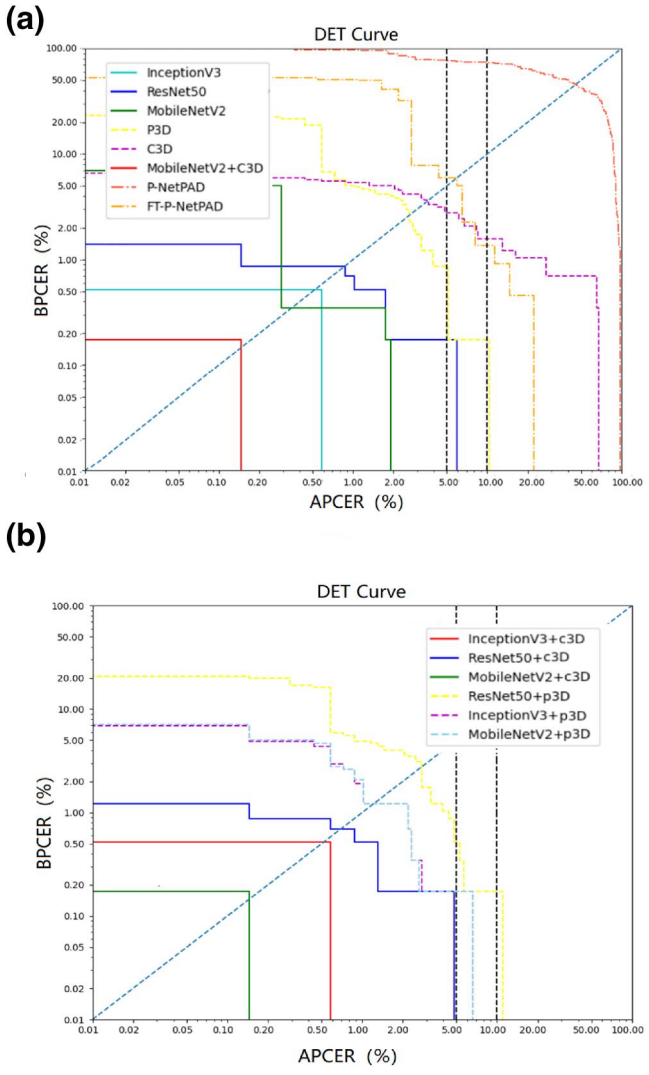

FIGURE 6 (a) Detection Error Trade-off (DET) curve of the best-performing variant, that is, *MobileNetV2 + C3D* and other iris presentation attack detection (PAD) methods. (b) DET curve of the variants of the proposed framework

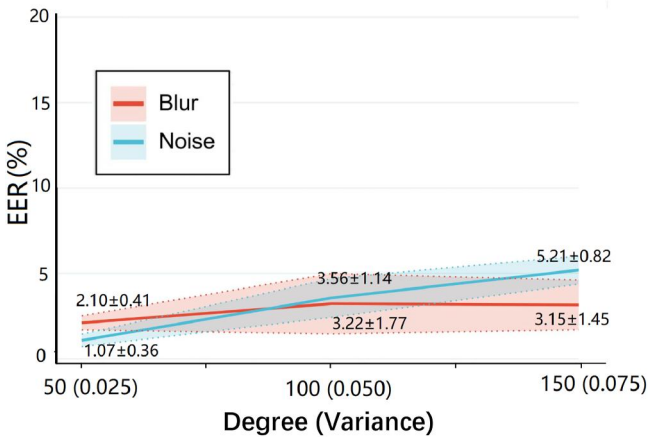

FIGURE 8 The performance fluctuation of the proposed framework against motion blur and Gaussian noise on the adopted dataset under 5-fold cross validation

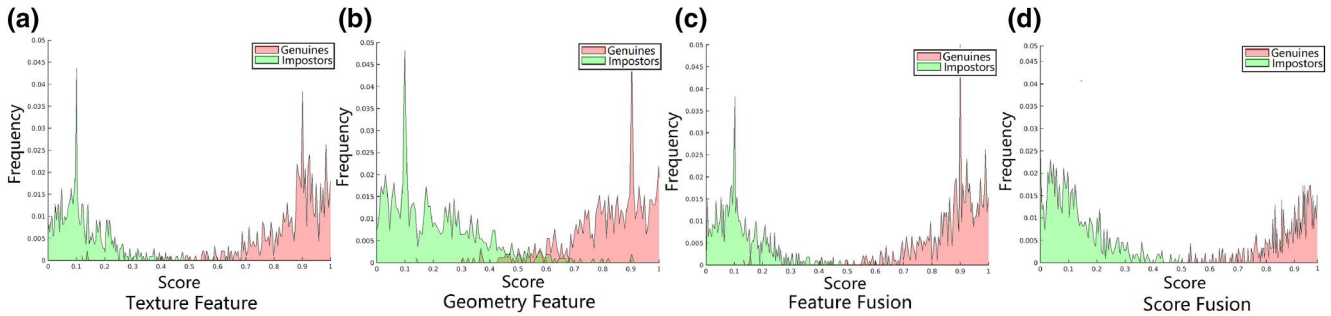

FIGURE 7 The score distribution of bona fide (genuines) and spoofing (impostors) iris images output by *ResNet50 + C3D*. Texture feature means the output of *ResNet50* and geometry feature means the output of *C3D*. Feature fusion means the output of *ResNet50 + C3D*, which combines the two branches at feature level. Score fusion means the average of the outputs of *ResNet50* and *C3D*, which fuses the two branches at score level

TABLE 3 The results of multi-class attack classification experiments

PAI	$BPCER_{10}$	$BPCER_{20}$	EER
Printed papers	0.00	1.39	3.33
Printed photos	0.00	0.17	0.82
Electronic display	8.34	12.87	8.66
Mean	2.78	4.81	4.27

As shown in Table 3, unknown PAIs can also be accurately detected by *MobileNetV2 + C3D* with a mean EER of 4.27%. The performance of *MobileNetV2 + C3D* is overall satisfactory, only slightly worse in detecting unseen attack of electronic display. The good generalisation ability on unseen presentation attacks is quite necessary for real-world applications.

4 | CONCLUSION

In the proposed framework, 2D texture and 3D geometric structure are extracted from LF focal stack and fused in a two-branch manner for iris PAD. To overcome the problem of LF data limitation, the off-the-shelf CNN features and optimised SVM classifier are integrated to combine the merits of LF imaging and DL frameworks for iris liveness detection. A group of pre-trained DL models oriented for dealing with 2D planar images and 3D sequences are adopted as feature extractor. Experimental results demonstrate that variants of the proposed framework outperform recent iris PAD methods that take 2D planar images or LF focal stacks as input in terms of standard evaluation metrics. The best-performing variant even beats fine-tuned SOTA models. The ablation experiments also verify that the fusion of 3D geometry and 2D texture features can achieve better detection performance than either of the single branches. In addition, the proposed framework exhibits good generalisation ability on unseen attacks and has considerable robustness against degradation factors such as motion blur and noise. The proposed framework may be embedded into an iris recognition system only equipped with an LF camera. It can be deployed in a resource-efficient manner without retraining bulky CNNs and collecting massive data on the spot. In future work, we will collect large-scale LF image databases with the lab-produced LF camera and consider to develop DL models tailored for iris PAD task.

ACKNOWLEDGEMENTS

This work is supported by the National Natural Science Foundation of China (Grant No. 62006225, 62176025, and 61906199), the Strategic Priority Research Program of Chinese Academy of Sciences (Grant No. XDA27040700) and sponsored by CAAI Huawei MindSpore Open Fund.

CONFLICT OF INTEREST

The authors declare no conflict of interest.

PERMISSION TO REPRODUCE MATRERIALS FROM OTHER SOURCES

None.

DATA AVAILABILITY STATEMENT

1. We have released the LF focal stack data on our website (http://www.cripacsir.cn/dataset/).
2. To reproduce the results in this paper, the source codes have been released on GitHub (https://github.com/luozhengquan/LFLD).

ORCID

Zhengquan Luo https://orcid.org/0000-0002-6348-3973
Yunlong Wang https://orcid.org/0000-0002-3535-308X

REFERENCES

1. Daugman, J.: Recognizing people by their iris patterns. Inf. Secur. Tech. Rep. 3(1), 33–39 (1998). https://doi.org/10.1016/s1363-4127(98)80016-2
2. Galbally, J., et al.: Iris liveness detection based on quality related features. In: 2012 5th IAPR International Conference on Biometrics (ICB), pp. 271–276. IEEE (2012)
3. He, Z., et al.: Efficient iris spoof detection via boosted local binary patterns. In: International Conference on Biometrics, pp. 1080–1090. Springer (2009)
4. Komulainen, J., Hadid, A., Pietikäinen, M.: Generalized textured contact lens detection by extracting bsif description from cartesian iris images. In: IEEE International Joint Conference on Biometrics, pp. 1–7. IEEE (2014)
5. Raghavendra, R., Busch, C.: Robust scheme for iris presentation attack detection using multiscale binarized statistical image features. IEEE Trans. Inf. Forensics Secur. 10(4), 703–715 (2015). https://doi.org/10.1109/tifs.2015.2400393
6. Alonso-Fernandez, F., Bigun, J.: Exploting periocular and rgb information in fake iris detection. In: 2014 37th International Convention on Information and Communication Technology, Electronics and Microelectronics (MIPRO), pp. 1354–1359. IEEE (2014)
7. Hu, Y., Sirlantzis, K., Howells, G.: Iris liveness detection using regional features. Pattern Recogn. Lett. 82, 242–250 (2016). https://doi.org/10.1016/j.patrec.2015.10.010
8. Kohli, N., et al.: Detecting medley of iris spoofing attacks using desist. In: 2016 IEEE 8th International Conference on Biometrics Theory, Applications and Systems (BTAS), pp. 1–6. IEEE (2016)
9. Das, P., et al.: Iris liveness detection competition (livdet-iris)-the 2020 edition. In: 2020 IEEE International Joint Conference on Biometrics (IJCB), pp. 1–9. IEEE (2020)
10. Zhang, K., et al.: Joint face detection and alignment using multitask cascaded convolutional networks. IEEE Signal Process. Lett. 23(10), 1499–1503 (2016). https://doi.org/10.1109/lsp.2016.2603342
11. Chen, C., Ross, A.: A multi-task convolutional neural network for joint iris detection and presentation attack detection. In: 2018 IEEE Winter Applications of Computer Vision Workshops (WACVW), pp. 44–51. IEEE (2018)
12. Sharma, R., D-netpad, A.R.: An explainable and interpretable iris presentation attack detector. In: 2020 IEEE International Joint Conference On Biometrics (IJCB), Pages 1–10. IEEE (2020)
13. Huang, G., et al.: Densely connected convolutional networks. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp. 4700–4708. (2017)
14. Juan, E., Gonzalez, S., Busch, C.: Tapia, Sebastian Gonzalez, and Christoph Busch. Iris liveness detection using a cascade of dedicated deep learning networks. IEEE Trans. Inf. Forensics Secur. 17, 42–52 (2022). https://doi.org/10.1109/TIFS.2021.3132582
15. Sandler, M., et al.: Mobilenetv2: inverted residuals and linear bottlenecks. In: 2018 IEEE/CVF Conference on Computer Vision and Pattern

Recognition, pp. 4510–4520 (2018). https://doi.org/10.1109/CVPR.2018.00474
16. SungLee, J., KangPark, R., Kim, J.: Robust fake iris detection based on variation of the reflectance ratio between the iris and the sclera. In: 2006 Biometrics Symposium: Special Session on Research at the Biometric Consortium Conference, pp. 1–6. IEEE (2006)
17. Adam, C.: Pupil dynamics for iris liveness detection. IEEE Trans. Inf. Forensics Secur. 10(4), 726–735 (2015). https://doi.org/10.1109/tifs.2015.2398815
18. 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. IEEE (2014)
19. Kim, S., Ban, Y., Lee, S.: Face liveness detection using a light field camera. Sensors 14(12), 22471–22499 (2014). https://doi.org/10.3390/s141222471
20. Raghavendra, R., Raja, K.B., Busch, C.: Presentation attack detection for face recognition using light field camera. IEEE Trans. Image Process. 24(3), 1060–1075 (2015). https://doi.org/10.1109/tip.2015.2395951
21. 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. IEEE (2016)
22. Sepas-Moghaddam, A., et al.: Face spoofing detection using a light field imaging framework. IET Biom. 7(1), 39–48 (2017). https://doi.org/10.1049/iet-bmt.2017.0095
23. Xie, X., et al.: One-snapshot face anti-spoofing using a light field camera. In: Chinese Conference on Biometric Recognition, pp. 108–117. Springer (2017)
24. Chiesa, V., Dugelay, J.-L.: Advanced face presentation attack detection on light field database. In: 2018 International Conference of the Biometrics Special Interest Group (BIOSIG), pp. 1–4. IEEE (2018)
25. Sepas-Moghaddam, A., Pereira, F., Correia, P.L.: Light field-based face presentation attack detection: reviewing, benchmarking and one step further. IEEE Trans. Inf. Forensics Secur. 13(7), 1696–1709 (2018). https://doi.org/10.1109/tifs.2018.2799427
26. Liu, M., et al.: Light field-based face liveness detection with convolutional neural networks. J. Electron. Imag. 28(1), 013003 (2019). https://doi.org/10.1117/1.jei.28.1.013003
27. Song, P., et al.: Iris liveness detection based on light field imaging. Acta Autom. Sin. 45(9), 1701–1712 (2019)
28. Tran, Du, et al.: Learning spatiotemporal features with 3d convolutional networks. In: Proceedings of the IEEE International Conference on Computer Vision, pp. 4489–4497. (2015)
29. Karpathy, A., et al.: Large-scale video classification with convolutional neural networks. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp. 1725–1732. (2014)
30. Qiu, Z., Yao, T., Mei, T.: Learning spatio-temporal representation with pseudo-3d residual networks. In: Proceedings of the IEEE International Conference on Computer Vision, pp. 5533–5541. (2017)
31. Kay, W., et al.: The Kinetics Human Action Video Dataset (2017). arXiv preprint arXiv:1705.06950
32. He, K., et al.: Deep residual learning for image recognition. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp. 770–778. (2016)
33. Szegedy, C., et al.: Rethinking the inception architecture for computer vision. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp. 2818–2826. (2016)
34. Jia, D., et al.: Imagenet: a large-scale hierarchical image database. In: 2009 IEEE Conference on Computer Vision and Pattern Recognition, pp. 248–255. IEEE (2009)
35. Ng, R., et al.: Light field photography with a hand-held plenoptic camera. Comput. Sci. Tech. Rep. CSTR 2(11), 1–11 (2005)
36. Pech-Pacheco, J.L., et al.: Diatom autofocusing in brightfield microscopy: a comparative study. In: Proceedings 15th International Conference on Pattern Recognition. ICPR-2000, vol. 3, pp. 314–317. IEEE (2000)
37. Dansereau, D.G., Pizarro, O., Williams, S.B.: Decoding, calibration and rectification for lenselet-based plenoptic cameras. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp. 1027–1034. (2013)
38. Information Technology-Presentation Attack Detection—Part 3: Testing, Reporting and Classification Ofattacks. Standard ISO/IEC 30107- 3:2017, International Organization for Standardization, September 2017. https://www.iso.org/standard/67381.html
39. Pedregosa, F., et al.: Scikit-learn: machine learning in Python. J. Mach. Learn. Res. 12, 2825–2830 (2011)
40. Moorthy, A.K., Bovik, A.C.: Blind image quality assessment: from natural scene statistics to perceptual quality. IEEE Trans. Image Process 20(12), 3350–3364 (2011). https://doi.org/10.1109/tip.2011.2147325
41. Ojansivu, V., Rahtu, E., Heikkila, J.: Rotation invariant local phase quantization for blur insensitive texture analysis. In: 2008 19th International Conference on Pattern Recognition, pp. 1–4. IEEE (2008)

How to cite this article: Luo, Z., et al.: Combining 2D texture and 3D geometry features for Reliable iris presentation attack detection using light field focal stack. IET Biome. 11(5), 420–429 (2022). https://doi.org/10.1049/bme2.12092