Face Liveness Detection Using a Flash Against 2D Spoofing Attack

Patrick P. K. Chan, *Member, IEEE*, Weiwen Liu, Danni Chen, Daniel S. Yeung, *Fellow, IEEE*, Fei Zhang, Xizhao Wang, *Fellow, IEEE*, and Chien-Chang Hsu, *Member, IEEE*

***Abstract*— Face recognition technique has been widely applied to personal identification systems due to its satisfying performance. However, its security may be a crucial issue, since many studies have shown that face recognition systems may be vulnerable in an adversarial environment, in which an adversary can camouflage as a legitimate user in order to mislead the system. Although face liveness detection methods have been proposed to distinguish real and fake faces, they are either time-consuming, costly, or sensitive to noise and illumination. This paper proposes a face liveness detection method with flash against 2D spoofing attack. Flash not only can enhance the differentiation between legitimate and illegitimate users, but it also reduces the influence of environmental factors. Two images are taken from a subject, one with flash and another without flash. Four texture and 2D structure descriptors with low computational complexity are used to capture information of the two images in our model. Advantages of our method include low installation cost of flash and no user cooperation required. A data set of 50 subjects collected under different scenarios is used in the experiments to evaluate the proposed method. The experimental results indicate that the proposed model performs better than existing liveness detection methods in different environmental scenarios. This paper confirms that the use of flash successfully improves face liveness detection in terms of accuracy, robustness, and running time.**

***Index Terms*— Face liveness detection, 2D spoofing attack, flash light, adversarial learning.**

I. Introduction

BIOMETRIC technology has been used widely in personal identification applications. As compared with the traditional security methods like passcodes, biometric technology brings about convenience which uses human intrinsic characteristics for individual identification [1], [2]. Face recognition is one of the most common biometric features because information from the face can be extracted easily without any physical contact. It has been successfully demonstrated in many personal identification applications, *e.g.* law enforcement, surveillance, information security, smart card authentication and entertainment [3]–[7].

Since traditional face recognition systems do not consider the existence of an adversary, many studies have revealed that these systems are vulnerable to spoofing attacks [8]–[10] in which an attacker obtains an illegitimate access to a system by camouflaging as an authorized person. A well-known example is a *2D spoofing attack*, which misleads a system by using a 2D facial duplicate of a valid user. As an image or a video of a person is easily obtainable and highly reproducible [11], [12], 2D spoofing attack is one of the most common attacks. There are three types of 2D spoofing attacks, namely photo attack, video attack and mimic mask attack. Photo attack evades the detection by using a picture of a legitimate user on a piece of paper [13], [14], or an electronic screen [15], while video attack misleads the system by using a video of an authorized person on electronic devices [16], [17]. In mimic mask attack, an adversary camouflages as an authorized person by wearing a 2D mask [18].

Face liveness detection [19], which is also referred to *face spoofing detection*, has been devised to defend against 2D spoofing attack. Face liveness detection determines whether an image is taken from a real or fake subject before face recognition process starts. Suspected images are filtered and will not be passed to the recognition system.

Previous works on face liveness detection mainly focus on *software-based methods* which analyze liveness clues, including texture [20], [21], structure information [22], [23] and liveness sign [24], of the subjects, and quality of captured images [15], [25], [26]. These methods are generally sensitive to environmental factors [19], [27], for instance, bad illumination condition and noisy images. Thus, their detection accuracy decreases significantly under such circumstances. In addition, computational complexity of calculating some liveness clue is high, *e.g.* facial dynamic is calculated based on consecutive frames [28]. Although asking users to speak [29] or shake their heads [30] improves the accuracy of the detection, it also reduces efficiency due to longer detection duration and uncooperative users. On the other hand, a device is embedded in a recognition system in *hardware-based methods* [31], [32] to capture additional information of the

Manuscript received November 23, 2016; revised March 16, 2017, June 27, 2017, and August 18, 2017; accepted September 17, 2017. This work was supported in part by Fundamental Research Funds for Central Universities under Grant 2015ZZ092 and in part by the National Training Program of Innovation and Entrepreneurship of China under Grant 201510561067. The associate editor coordinating the review of this manuscript and approving it for publication was Prof. Domingo Mery. *(Corresponding author: Fei Zhang.)*

AQ:1 P. P. K. Chan and D. Chen are with the School of Computer Science and Engineering, South China University of Technology, Guangzhou, China (e-mail: patrickchan@ieee.org; conniechen9469@gmail.com).

W. Liu is with the Department of Computer Science and Engineering, The Chinese University of Hong Kong, Hong Kong (e-mail: patrickchan@ieee.org).

AQ:2 D. S. Yeung is with ???

F. Zhang is with the College of Computer and Information Engineering, Henan Normal University, Xinxiang, China (e-mail: zhangfei@htu.edu.cn).

X. Wang is with the College of Computer Science and Software Engineering, Shenzhen University, Shenzhen, China (e-mail: xizhaowang@ieee.org).

C.-C. Hsu is with the Computer Science and Information Engineering, Fu Jen Catholic University, Taipei, Taiwan (e-mail: cch@csie.fju.edu.tw).

Color versions of one or more of the figures in this paper are available online at http://ieeexplore.ieee.org.

Digital Object Identifier 10.1109/TIFS.2017.2758748

TABLE I
SUMMARY OF EXISTING METHODS AGAINST 2D SPOOFING ATTACK

Category	Sub-category	Description	Typical Algorithms	Pros	Cons
Software-based	Texture	Capture difference on visual and tactile quality between real and fake faces	local binary patterns(LBP) [34], Fourier analysis [20], color texture analysis [35], etc	Low implementation cost and low time complexity	Easily affected by illumination condition, noise and image quality
	Structure Information	Capture difference of structure properties between 3D real faces and 2D-planar attack	diffusion speed [18], facial feature trajectories [23], defocusing techniques [18], optical flow [22], [36], [37], etc	Relatively high detection accuracy	High time complexity, sensitive to illumination and image quality
	Liveness Sign	Capture natural human movements	Detection of eye blinking [24], [38], [39], head rotation [30] and lip movements [40]	Performs well in attacks with no human dynamics, like photo attack and mask attack	Fail to evade video attack, long detection time, high space and time complexity
	Image Quality Analysis	Analyze the quality of the real face and 2D spoof face images	Analysis of image specularity distribution [25], image distortion [15], [41] and general features [26]	Good generalization ability to various scenarios	Device dependent; Attack media with high resolution may fool the detection system
	Hybrid Methods	Combine different kinds of information to assist the detection	DMD-LBP-SVM, which combines texture and structure information [28]	Substantial information makes the detection more accurate	Longer time for feature processing leads to low detection efficiency
Hardware-based		Use additional hardware to measure the properties of a live face, like temperature and the reflectance of the subject	Infared camera [42], 3D camera, multiple 2D cameras [43] , light field camera [44], etc	High detection accuracy	High setup and maintenance cost

subjects, *e.g.* temperature. Nevertheless, some of the additional hardware is costly and difficult to install. Our preliminary study [33], which only analyzes the difference of the hair on foreheads between real and fake faces, showed that flash increases the differentiation between a legitimate person and the 2D spoofing attack. However, the study only focused on video attack in a particular environmental setting in which the ambient illumination is normal, and the distance between the camera and the background is short. The usefulness of flash on detecting other 2D spoofing attacks remains unclear. Moreover, the proposed model is sensitive to the hair on the forehead and may not be practical since users have different hair styles. Therefore in this paper we provide a complete investigation on how the use of flash can improve 2D spoofing attack detection. The literature review of face liveness detection and also 2D spoofing attack is introduced in Section II.

In Section III, a model of face liveness detection using flash to defend against photo, video and also mimic mask attacks will be elaborated. In the proposed model, a pair of images is taken from a subject in the detection, one with flash and the other without flash. Features of our method are carefully designed in order to provide accurate and robust prediction with low time complexity. The descriptor based on uniform local binary patterns is applied to measure the textural information from the face, and another three descriptors are proposed to capture the structure information of a face using the standard deviation and the mean of grayscale difference between the images with and without flash.

Then, the subject is classified as either legitimate or malicious class based on the difference between the images with and without flash measured by the four descriptors. Unlike hardware-based methods, our method requires only flash which is economical and easy to install in existing face recognition systems. The proposed method is expected to be more accurate and robust than the software-based method since flash enhances the differentiation between real and fake faces and reduces the influence of ambient illumination. In addition, the time complexity of extracting the four descriptors is low and no user cooperation is required. Our method takes advantage of both software and hardware based methods. The discussion on the reasons why considering the difference between the images with and without flash is helpful in face liveness detection based on the Lambertian reflectance law is also provided.

In Section IV, the performance of the proposed model is then evaluated and compared with other well-known face liveness detection methods under different environmental settings, including background distance and ambient illumination. The procedure of the dataset collection is also described. Finally, the conclusion and future work are given in Section V.

II. LITERATURE REVIEW

Existing face liveness detection methods against the 2D spoofing attack are briefly introduced in this section. According to the requirement of an additional device, face liveness detection methods can be categorized into software-based and hardware-based method respectively. The pros and cons in accuracy, time complexity, implementation cost and convenience to users will also be discussed. Table I summarizes the existing 2D spoofing attack detection methods.

Software-based method is the most widely used face liveness detection method. It determines whether a target is of the real face based on the information of the captured images, that is, the texture, structure information, liveness sign and image quality, without using additional hardware device. The light reflection of real human skin is different from the one displayed on a 2D-planar object, *i.e.* a paper or a mobile, in 2D spoofing attack. This difference in the visual and tactile quality is captured by *texture-based methods*. The well-known example is local binary patterns (LBP) [34] which labels the pixels of an image by thresholding the neighborhood of each pixel to represent the local texture information with the property of invariance to monotonic grayscale transformation. Generally, an image can be divided into several blocks, and LBP histograms are extracted individually. For each block, the LBP code of a pixel (x_c, y_c) is calculated using bilinearly interpolating values at non-integer sampling points in its neighborhood, as shown in (1).

$$LBP_{P,R}(x_c, y_c) = \sum_{i=0}^{P-1} g(p_i - p_c) \times 2^i, \tag{1}$$

where p_c is the gray value of the pixel (x_c, y_c) and p_i refers to the gray value of the i^{th} pixel. P and R are parameters of LBP, which represent P sampling points on a clockwise circle of radius R for each pixel's neighborhood. The function $g(z)$ is a threshold function, which outputs 1 when z is non-negative; otherwise, outputs 0. The occurrences of LBP codes are represented by a histogram. The numbers of occurrence are applied as input vectors for training.

The advanced LBP feature, referred to uniform LBP feature [34] ($LBP_{P,R}^{u2}$), is also proposed to reduce the dimensionality of the original LBP feature, which has been widely adopted in face liveness detection recently. An LBP code is uniform if it contains at most two bitwise transitions from 0 to 1 or vice versa. Each uniform LBP code is considered individually, and the rest of the non-uniform ones are grouped into one bin in the histogram. As a result, time complexity is significantly reduced since the non-uniform LBP codes are ignored. Another example of texture-based methods is the color texture of analyzing both luminance and chrominance channels which also exhibit effectiveness in 2D spoofing detection [35]. Difference of Gaussians (DoG) [14], which is a bandpass filter considering two Gaussian functions with different variances, has also been applied to improve the accuracy of the face liveness detection by removing the variant lighting in a face image. Fourier analysis [20] measures the frequency domain of face images, which is another texture information. The major drawback of a texture-based method is that its performance is highly affected by illumination condition and the quality of the input image [27]. Although the implementation cost and the time complexity are relatively low, some unexpected factors like uneven illumination and camera noise can degrade the performance significantly.

Structure information, which reveals information of the 3D structure of a subject from the projected 2D image, is also used in some detection methods. Illumination of 2D surface diffuses more slowly than that of 3D since its intensity is more evenly distributed. Diffusion is measured by the features of local speed patterns for the Diffusion Speed method (DS) [18] in order to detect a live face. Thus it is faster due to non-uniformity of the 3D surface. In addition, the depth of a face is analyzed by the facial feature trajectories [23] and the defocusing technique [18], which is a common technique for structure information. Several works on different movement patterns of 2D planes and 3D objects by optical flow fields are also captured [22], [36], [37]. The major drawbacks of these methods are high time complexity, sensitivity to the illumination and the quality of the images [36].

Some studies which focus on *liveness sign*, usually refer to the natural human movements. For example, eye blinking [24], [38], [39], head rotation [30] and lip movement [40] are common ones. Obviously, methods of this kind are designed specifically for image attacks. However, video attack is able to evade these methods easily [45], [46]. Moreover, a video has to be stored in order to detect a particular movement. This kind of method usually requires a longer detection time, and also larger space and computational complexity.

The quality of a face image in a 2D spooking attack may degrade since the face image is obtained by recapturing from photos and videos. *Image quality* has been used as an indicator in face liveness detection. For instance, the difference of specularity spatial distribution between a recaptured image and its original image [25], the distortion of a spoof attack image with respect to specular reflection, blurriness, chromatic moment, and color diversity [41], and the image quality based on 25 metrics [26] are studied. High Definition (HD) camera and display increase the resolution of mimic, which may increase the difficulty of detection by image quality analysis.

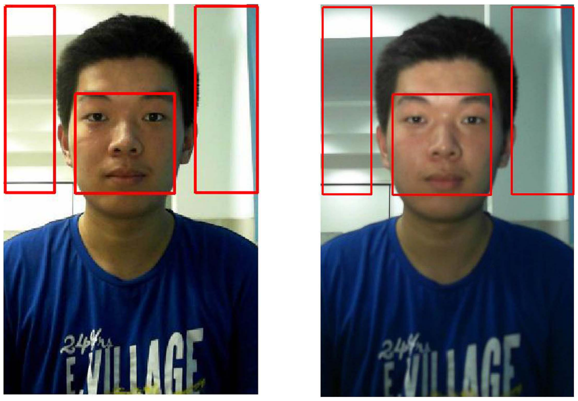
Some methods are also proposed by using different kinds of features in order to achieve higher accuracy. For instance, the features of liveness sign and texture of sequential image frames are used in dynamic mode decomposition (DMD) [28]. The model applies eye blinking, lip motion, facial expression change as well as LBP features to distinguish legitimate users from 2D spoofing attack. Another example is to apply eye blinking and background context texture to detect spoofing attack [45]. Although the time complexity is higher, the detection is usually more accurate.

In contrast, *hardware-based methods* require extra hardware to measure the additional information of subjects other than the camera of the face recognition system. A thermal camera, which has been successfully applied to face recognition [47], captures temperature and reflectance distribution of a subject. The Intensity and Texture Encoder (ITE) features [42] containing LBP and intensity histogram to detect non-biometric patches are extracted from a thermal image; a 3D camera or multiple 2D cameras [43] can be used to generate the 3D model of the subject; and a light field camera captures the light distribution of the subject [44]. Although hardware-based methods usually outperform software-based methods, the setup cost of extra devices is also much higher [1], [3].

Some detection methods need the cooperation of users. The users have to complete certain tasks during the detection process. For example, the user is required to speak for the audio-visual matching process [29], [48], [49], and to rotate the head for the 3D structure recovering process [50]. These methods achieve more accurate results at the cost of user inconvenience. However, the detection time needed is normally longer than that without user cooperation requirement.

III. Liveness Detection Method Based on Flash and No Flash Image Pairs

The proposed liveness detection method which takes advantages of both software and hardware based methods is



(a) (b)

Fig. 1. Examples of result of the face and the background extraction. The center rectangle and the rectangles on both sides of each image are the face and the background region: (a) Non-flash image; (b) Flash image.

Algorithm 1 Procedure of Feature Extraction of the Proposed Model

Input: I_n: the non-flash image; I_f: the flash image
Output: LBP_FI, SD_FIC, M_BIC and SD_BIC descriptors

1: identify I_n^F and I_n^{BG} from I_n based on a face identification method;
2: identify I_f^F and I_f^{BG} according to the locations of I_n^F and I_n^{BG} respectively;
3: extract descriptor LBP_FI from I_n^F;
4: $D^F = I_f^F - I_n^F$;
5: descriptor SD_FIC = std(D^F);
6: calculate $D^{BG} = I_f^{BG} - I_n^{BG}$;
7: descriptor M_BIC = mean(D^{BG});
8: descriptor SD_BIC = std(D^{BG}).

introduced in this section. An additional device, flash, is applied to enhance the performance of the software based method which considers the texture analysis and the structure information. The underlying principle is to magnify the differences between real face and fake face displayed in 2D media by using flash.

During the detection, two images with and without flash, denoted as I_f and I_n, are taken for the subject. We identify the rectangle regions for the face and the background defined by the pixels in the upper right corner and in the lower left corner of the region in I_n. The face region I_n^F is firstly determined. We apply the split up Sparse Network of Winnows (SNoW) classifier [51], one of the efficient face identification methods based on Successive Mean Quantization Transform. Two background regions, denoted as I_n^{BG}, are therefore located based on the face region. Specifically, the upper right corner and the lower left corner of the rectangle region of the right I_n^{BG} are defined by the upper right corner of I_n and 20 pixels to the right of the right corner of I_n^F to avoid the hair of a subject being selected. The left I_n^{BG} is defined similarly. Finally, I_f^F and I_f^{BG} are extracted from I_f according to the locations of I_n^F and I_n^{BG} respectively. Examples of the result of the face and background extraction are shown in figure 1.

Four carefully designed descriptors including LBP_FI, SD_FIC, M_BIC and SD_BIC are extracted from both regions of the face and the background. These descriptors should be able to distinguish legitimate users and the common 2D spoofing attack efficiently, accurately and robustly. The photo attack printed on a paper, the photo attack displayed on iPad, the video attack, the 2D mask attack and the curved mask attack are considered. The curved mask attack is considered as an extension of a 2D attack since it misleads the recognition system by holding the 2D mask curly. It is more difficult to detect the curved mask attack than the 2D mask attack since the curved mask covers the face more tightly than the 2D mask attack. The descriptors are input as features to a classifier for detection. The procedure for feature extraction of the proposed model is described in Algorithm 1. A real face can be distinguished from a fake one by a classifier using the extracted features. Support Vector Machine (SVM) is used in our model due to its simplicity and satisfying performance in a two-class classification problem.

In this section, the four descriptors are firstly introduced in Section III-A. Then, the underlying rationale of the proposed model is discussed in Section III-B.

A. Descriptors of the Model

1) Uniform Local Binary Patterns on the Flash Image (LBP_FI) Descriptor: LBP analysis is applied to capture the local texture information of the face region of the image with the flash (I_f^F). The reason of using I_f^F only is that the flash increases the detail of the real face but not the fake one due to the difference between 3D and 2D surfaces. As a result, a legitimate user can be distinguished from the camouflaged one.

I_f^F is firstly separated into nine non-overlapping blocks to obtain the texture information from different regions of the image [21]. The LBP code of the pixel (x, y) in each block is then calculated. In our model, the circle of radius is set to 1 and all neighbor pixels are considered, *i.e.* $P = 8$ and $R = 1$.

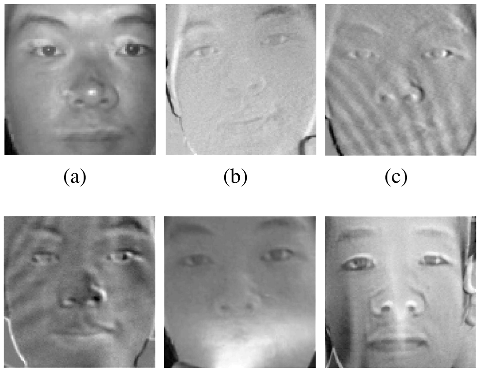
Since it has been shown that the uniform LBPs account for a bit less than 90% of all patterns in this setting [52], (1) of the LBP code can be simplified as (2).

$$LBP(x_c, y_c) = \sum_{i=0}^{7} g(p_i - p_c) \times 2^i. \tag{2}$$

There are totally 59 bins including 58 uniform patterns and the one containing the rest of the non-uniform patterns. The histogram $\mathbf{H}_i$ is generated according to $LBP(x_c, y_c)$ for the i^{th} block, where $\mathbf{H}_i = (h_1, h_2, \ldots, h_{59})$ and h_j is the occurrence of a pattern in j^{th} bin. Subsequently, there are a total of 531 (*i.e.* 9×59) values in LBP_FI, as shown in (3).

$$\text{LBP\_FI} = (\mathbf{H}_1, \mathbf{H}_2, \cdots, \mathbf{H}_9) = (h_1, h_2, \cdots, h_{531}). \tag{3}$$

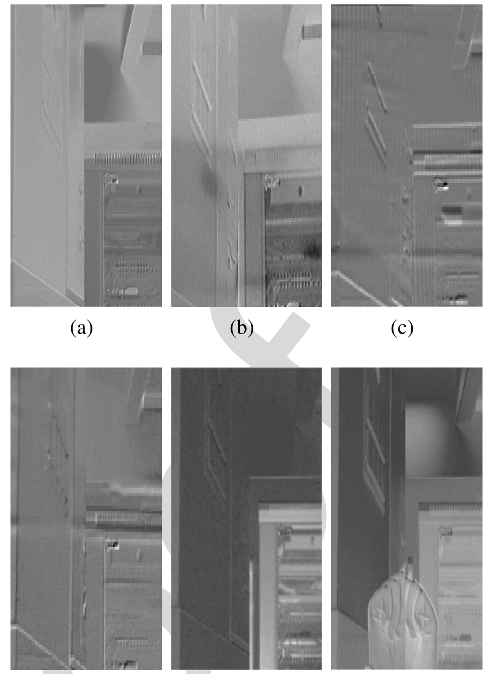
2) Standard Deviation of Face Intensity Change (SD_FIC) Descriptor: SD_FIC measures the grayscale intensity change of the face region caused by flash. The reflection of flash varies in the real face due to its structure information, *i.e.* the



(a) (b) (c)

(d) (e) (f)

Fig. 2. Examples of the face difference images for real face and different types of attacks: (a) Real face: SD_FIC=39.45; (b) Paper photo attack: SD_FIC=19.42; (c) iPad photo attack: SD_FIC=18.52; (d) Video attack: SD_FIC=17.03; (e) 2D mask attack: SD_FIC=30.44; (f) Curved mask attack: SD_FIC=33.80.



(a) (b) (c)

(d) (e) (f)

Fig. 3. Examples of the background difference images for real face and different types of attacks: (a) Real face: M_BIC=36.88, SD_BIC=24.02; (b) Paper photo attack: M_BIC=62.12, SD_BIC=25.81; (c) iPad photo attack: M_BIC=58.87, SD_BIC=17.13; (d) Video attack: M_BIC=63.24, SD_BIC=13.11; (e) 2D mask attack: M_BIC=35.57, SD_BIC=37.76; (f) Curved mask attack: M_BIC=43.88, SD_BIC=33.88.

distances between the flash and each part of the face may be different. In contrast, the reflected light of a 2D spoofing attack is more uniform. As a result, the deviation of the intensity of the real person is larger than that of a 2D spoofing attack. The standard deviation is applied to capture the change of the grayscale intensity in our model, and SD_FIC is defined as in (4).

$$\mathrm{SD\_FIC} = \sigma_{D^F} = \sqrt{\frac{\sum_{i=1}^{N}(D^F(x_i, y_i) - \mu_{D^F})^2}{N-1}}, \qquad (4)$$

where μ_{D^F} and σ_{D^F} denote the mean and the standard deviation of $D^F(x, y)$ respectively, N is the number of pixels in the region and $D^F(x, y) = I_f^F(x, y) - I_n^F(x, y)$. The reason for deducting the intensity of the image without the flash light in $D^F(x, y)$ is to reduce the influence to the ambient illumination. The examples of D^F of the real face and the different types of attacks, as well as their SD_FIC values, are shown in figure 2. As discussed, the value of SD_FIC of the real face is the largest among all cases due to the intensity change on the 3D object. The paper photo, 2D mask and curved mask attacks have a larger SD_FIC than other types of attacks because a bright strip occurs in the face region.

3) Mean of Background Intensity Change (M_BIC) Descriptor: The actual background has been blocked in the photo and video attacks. As the captured background on the display media is much closer to the camera than the actual one, higher intensity of light will be reflected. We propose the M_BIC to capture this information, defined as follows:

$$\mathrm{M\_BIC} = \mu_{D^{BG}} = \frac{\sum_{i=1}^{N} D^{BG}(x_i, y_i)}{N}, \qquad (5)$$

where $D^{BG}(x, y) = I_f^{BG}(x, y) - I_n^{BG}(x, y)$, $-255 \leq D^{BG} \leq 255$ and $D^{BG} \in Z$. Examples of D^{BG} of the real face and the different types of attacks are illustrated in figure 3. D^{BG} is linearly mapped to a range of 0 to 255 in the illustration to avoid the negative value. Therefore, the darker area indicates I_n^{BG} is much larger than I_f^{BG}. As different from the real face and the two mask attacks, the real background is blocked in the image with flash for the photo and video attacks. The values of their D^{BG} are much larger than the ones without flash, *i.e.* their M_BIC values are larger. On the other hand, the real face and the two mask attacks have close M_BIC values because their backgrounds are real and the effect of flash on them is quite similar.

4) Standard Deviation of Background Intensity Change (SD_BIC) Descriptor: As different from the photo and video attacks mentioned in the previous section, the actual background is not covered since only the region of a subject's head is used in the 2D mask attack or curved mask. The light diffusion of masks is different from the one of real face due to the texture and the shape. The light intensity of I_f^{BG} of legitimate and malicious users is different. The variation of the light intensity is measured by

$$\mathrm{SD\_BIC} = \sigma_{D^{BG}} = \sqrt{\frac{\sum_{i=1}^{N}(D^{BG}(x, y) - \mu_{D^{BG}})^2}{N-1}}. \qquad (6)$$

Figure 3 shows SD_BIC values of the real face is smaller than that of the mask attacks. It is because the light diffusion of the mask is larger than that of the face. Moreover, the hands captured in the curved mask attack also increase its SD_BIC. Due to a 2D planar structure of an iPad and a photo, flash increases the intensity of the background region uniformly in iPad and paper photo attack, *i.e.* SD_BICs of these attacks are relatively smaller than the ones not covering the real background.

B. Conceptual Discussion

Assume $I(x, y)$ denotes the intensity or grayscale value of the pixel (x, y), where $I(x, y) \in \mathrm{Z}$ is in $[0, 255]$. The intensity of the image without flash (I_n) is defined in (7) according to the Lambertian reflectance law [53]

$$I_n(x, y) = K L_a, \tag{7}$$

where $K \in (0, 1)$ denotes a surface reflectivity at pixel (x, y). Larger K indicates more intensive light is reflected from the surface. $L_a \in (0, \infty)$ is the intensity of the ambient illumination. $L_a = 0$ indicates the dark environment. The model assumes only the ambient light is considered and the intensity of the ambient light is a constant at any point and direction. Therefore, without any additional lighting, as L_a is the same for any object in the same environment, only K is useful for the face liveness detection, *i.e.* the smoothness of a human skin and that of a fake one displayed on 2D planar material are different. However, a face liveness detection only considering K is sensitive to the quality of images and the change of illumination, which has been shown by experiments in the previous study [54].

Based on the Lambertian reflectance law, one additional component is added to the intensity of the image with flash (I_f) defined in (8). In order to make a difference between the scaler and vector multiplication, we omit the dot of the scaler multiplication in these two equations.

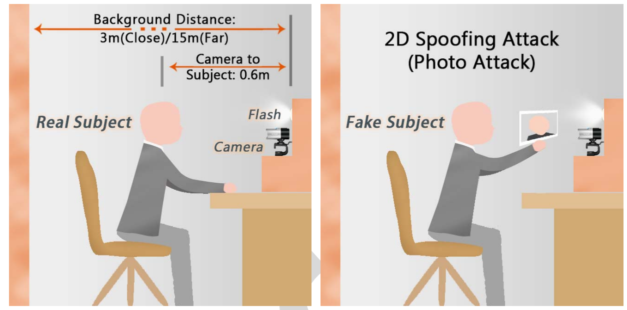
$$I_f(x, y) = K L_a + K L_f \frac{\mathbf{N} \cdot \mathbf{T}}{r^2} = K L_a + K L_f \frac{\cos\theta}{r^2}, \tag{8}$$

where $L_f \in (0, \infty)$ denotes the intensity of the flash. $\mathbf{N}$ is the normal vector to the object surface and $\mathbf{T}$ represents a normalized light-direction vector, pointing from the object surface to the source of flash. θ denotes the angle between $\mathbf{N}$ and $\mathbf{T}$, $\theta \in [0, 90°]$. r is the distance between the flash and the point of the surface. θ as well as r, and $I_f(x, y)$ are inversely proportional, *i.e.* larger θ or r decreases $I_f(x, y)$.

Under the same lighting condition (*i.e.* L_a and L_f are fixed), θ and r of subjects are different due to their shapes. As a result, not only the texture information but also the structure information will be measured. In our proposed model, the LBP_FI descriptor captures the texture information, while SD_FIC, M_BIC and SD_BIC measure the structure information. As a result, the second term of (8) provides extra information to separate the legitimate users from the 2D spoofing attack. It explains why our method may be more accurate than the ones without flash. In addition, more stable liveness detection is expected because of flash, which has a relatively strong illumination in comparison with the ambient light, and it reduces the influence of ambient illumination.

IV. Discussion on Experimental Results

In this section, the performance of our proposed face liveness detection method to encounter different 2D spoofing attacks is evaluated and compared with existing methods experimentally using the dataset we collected under different scenarios. The procedure of the dataset preparation is described at the beginning. Then, the experimental settings as well as the evaluation criterion are introduced. Finally, the experimental results are given and discussed.



(a) (b)

Fig. 4. Settings of sample collection for our dataset: (a) A real subject; (b) A fake subject under photo attack.

(a) (b) (c) (d)

Fig. 5. Examples of the collected images with different distances under normal and uneven ambient illuminance: (a) Far distance (15m) under normal illuminance; (b) Far distance (15m) under uneven illuminance; (c) Close distance (3m) under normal illuminance; (d) Close distance (3m) under uneven illuminance.

A. Dataset Collection

The dataset[1] for the face liveness detection containing 50 subjects is collected in this paper. The group of subjects consists of 42 male and 8 female with the age from 18 to 21. Each subject is required to sit in front of a web camera (*i.e.* Microsoft Lifecam Studio [55]). Two images, one with flash and another without flash, are taken within a second. Images with 240×360 px are captured, and the face region is around 100×100 px. The detailed setting of the sample collection is illustrated in figure 4.

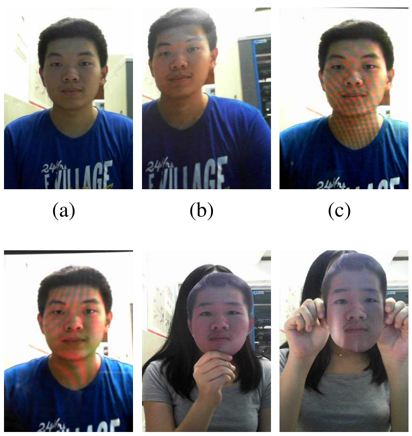
The distance between a subject and the camera is 0.6m. The flash is set up right above the camera. The distance between the subject and the background is set at 3m and 15m respectively in order to investigate how the distance to background affects the accuracy of liveness detection. The uneven illumination condition, *e.g.* the recognition system is next to a window, is also simulated. A lamp is placed by the side of the subject to create the unbalanced lighting environment. The images with different distances to the background and illumination conditions are shown in figure 5.

We use illuminance, defined as the total luminous flux incident on a surface per unit area, to represent the intensity of light. Illuminance measures how much incident light illuminates the surface. The only ambient light source in the room in the experiment is ceiling lighting. The illuminance meter is put on the top of the face of a subject, which is

[1] http://www.mlclab.org/dataset/FaceLiveFlash.htm

(a) (b) (c) (d) (e)

Fig. 6. Examples of collected images with additional illuminance values of the target: (a) No extra light; (b) +40lx; (c) +80lx; (d) +120lx; (e) +160lx.



(a) (b) (c) (d) (e) (f)

Fig. 7. Examples of real face and different types of attacks: (a) Real face; (b) Paper photo attack; (c) iPad photo attack; (d) Video attack; (e) 2D mask attack; (f) Curved mask attack.

parallel to the light source on the ceiling. Without additional device, the natural lighting of the subject is approximately equal to 40lx. To avoid the discomfort to human eyes, we limit the intensity of flash in our proposed method. Four different intensity levels of flash are set to increase the illuminance of the subject by +40lx, +80lx, +120lx, and +160lx. The maximum illuminance adopted by our method, which is 200lx (*i.e.* 40 + 160lx) at 0.6m, is much less than the flash for the camera. For example, the illuminance of the flash for Sony cameras HVL-F60M [56] and HVL-F43RM [57] are approximately 600lx and 400lx respectively at 0.5m. These ensure that the proposed method is practical and the intensity of flash is within the endurance of human eyes. Images with different illuminance values are illustrated in figure 6.

We simulate five different types of 2D spoofing attacks for each person: 1) the photo attack on the A4 sized photographic paper (paper photo attack), 2) the photo attack displayed on iPad with 1024 × 768 px screen (iPad photo attack), 3) a video (30 fps) being played on iPad with 1024 × 768 px screen (video attack), 4) the 2D mask attack with the background cut out (2D mask attack), and 5) the curved mask attack with the background cut out (curved mask attack). The examples of a real person and his/her 2D spoofing attacks are shown in figure 7.

For the legitimate user, 2D mask attack and curved mask attack, by considering the distance between the background and the subject, the ambient illumination, and flash illumination, 20 different photos are taken for each person. A total of 1000 samples are collected for each of these classes. Differently, for paper photo attack, iPad photo attack and video attack, the distance between the background and the subject is not considered since the real background cannot be captured. As a result, only 500 images are collected for each of them.

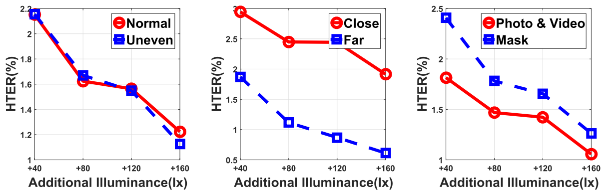
In addition, one thermal image method, which is a hardware based method, is also considered in the experiments. Additional thermal images are collected from 21 subjects by the thermal camera called Seek Thermal Compact XR [58] on a smartphone. The spectral range of the thermal camera is from 7.5 to 14 microns, with 206 × 156 px image resolution. The low-quality thermal camera is considered since its price is much lower than the professional ones. Therefore, it is more likely to be widely adopted in practice. The factors of environmental illumination and background distance are neglected since they do not affect the decision of a thermal image method. As a result, a total of 126 thermal images were taken, including 21 real face and 105 2D spoofing attack samples.

Temperature of a subject in the real face samples is 33 - 35 °C. As for a paper photo, which is used in 2D mask and curved mask attack, the temperature of a subject in these attacks is 28 - 30 °C, while the one in iPad photo attack is 30 - 32 °C. To evaluate the robustness of the thermal image method, the attack samples are camouflaged by increasing the temperature of 2D spoofing attack. A hot object (*i.e.* a heat pack) is put on the top of the papers, the iPads, and the masks used in the 2D spoofing attack before these objects are put in front of the camera, in order to increase the temperature by 2 - 4 °C. As a result, the temperature difference between a real face and the attack is reduced.

B. Experimental Setting and Evaluation Criterion

The experiments are performed on a computer with 8GB of memory and one Intel processor with i5-4210U cores at 2.40 GHz. A Support Vector Machine (SVM) with the Gaussian kernel implemented by libSVM [59] is applied as the classifier in the experiments. The parameter selection of the penalty coefficient C and Kernel radius γ follow the method of five-fold cross validation using training set based on grid search, which maximizes the classification accuracy. Six methods are selected from different categories of the existing face liveness detection to compare with our proposed method: 1) Traditional LBP method (LBP) [34] in texture-based methods, 2) Eye blinking detection method (EB) [24] in liveness-sign-based methods, 3) Optical Flow Field method (OFF) [22], 4) Diffusion Speed method (DS) [18] in 3D-structure-information-based methods, 5) DMD-LBP-SVM method (DLS) [28] in hybrid methods, and 6) thermal image (TI) in hardware-based methods. A preliminary evaluation is run to tune the parameters of all methods aiming to maximize their average accuracies.

For each experiment, the five-fold cross validation is applied. The performances of the liveness detection methods are evaluated by the running time and also a commonly used criterion, Half Total Error Rate (HTER). HTER is half of False Rejection Rate (FRR) and False Acceptance Rate (FAR), which are both determined by a threshold τ. FRR and FAR are monotonic increasing and decreasing



(a) (b) (c)

Fig. 8. The change of average HTER (%) of the proposed method under different settings and attack types: (a) Under normal and uneven illuminance; (b) Under close and far background distance; (c) Under photo & video and mask attacks.

functions of τ respectively. Larger τ indicates that there is a less probability that a spoof face is misclassified as a live one, and vice versa. When τ is set to the point where FRR and FAR are equal, HTER reaches its minimum. For a dataset $\mathcal{D}$, HTER is defined by

$$HTER(\tau, \mathcal{D}) = \frac{FRR(\tau, \mathcal{D}) + FAR(\tau, \mathcal{D})}{2}, \tag{9}$$

where the range of HTER is from 0 to 1. Lower HTER indicates that the system performs better.

C. Results and Discussion

In this section, we first discuss how the illuminance of the flash affects the performance of our method. Then the proposed model is compared with the existing methods in different scenarios, *i.e.* normal and uneven illumination, the distance between the subject and background, the quality of images and the computational complexity. The discriminate ability of descriptors used in our method is also evaluated. Finally, the performance of the proposed method with the partial knowledge on the type of attacks is discussed.

1) Proposed Method With Different Flash Light Illuminance: This section evaluates how the parameter, the additional illuminance value on the subject increased by flash, affects the performance of the proposed model in different environmental conditions. For each illuminance value and environmental setting, an SVM classifier is trained to distinguish the legitimate users from one type of 2D spoofing attacks. The average performance of the proposed model in different scenarios such as the normal and uneven ambient illuminance, close and far background distance, and photo & video and mask attacks are shown in figure 8. The x-axis and y-axis of the figures represent the additional illuminance values on the subject caused by flash and the average HTER respectively.

In all cases, the values of HTER of the proposed model decreases with the increase of the additional illuminance on the subject. There is no noticeable difference on the increase rates in normal and uneven ambient illumination since flash reduces the influence of the uneven ambient to the detection. However, as the difference between a subject and a background increases by flash, HTER drops more gently in the close distance scenario than the ones in the far distance scenario. As mentioned, detection on mask attacks is more difficult than photo and video attacks since the real background is not blocked by mask attacks. By increasing illuminance, more detail of a mask can be captured. This information is useful to distinguish a mask from a real face. That is why the improvement in the detection of the mask attacks is more significant than that of photo and video attacks.

The results suggest that using a flash light is useful to distinguish 2D spoofing attacks from the legitimate users. Moreover, flash with higher intensity improves the accuracy of the proposed model. This finding is consistent with our explanation of adding flash light in our model in Section III-B. On the other hand, strong flash light will cause the eyes of the users uncomfortable. This parameter is a trade-off between the effectiveness of the liveness detection system and its user friendliness. Two flash settings, *i.e.* +120lx and +160lx shown in figures 6d and 6e, are chosen for the comparison experiments in Sec IV-C.2 and Sect. IV-C.3 to illustrate the performance of our methods using different settings.

2) Comparison With Existing Methods Under Different Attacks: Our proposed methods with +120lx and +160lx, and the five software-based face liveness detection methods, including Traditional LBP method (LBP), Eye blinking detection method (EB), Optical Flow Field method (OFF), Diffusion Speed method (DS), DMD-LBP-SVM method (DLS), and one hardware-based method, *i.e.* thermal image (TI), are evaluated under the 2D spoofing attacks in different environmental settings.

The Student's t-test is conducted to evaluate the confidence level on the difference between the performance of our methods and others. The values of HTER of these experimental results are shown in Table II.

The experimental results indicate that the proposed method with +160lx has the lowest HTER under any type of attack. Moreover, most of the results show that the difference of our method with +160lx and others is statistically significant. On the other hand, our method with +120lx is slightly worse than the one with +160lx in general. These results are consistent with the previous section. Although a soft flash is used, the method with +120lx is still better than the comparison methods in most cases. The results suggest that the use of the flash light improves the 2D spoofing attack detection. The intensity of flash is an important parameter which significantly affects the accuracy of our method.

The proposed method with +160lx is statistically more significant than others in normal illumination with 95% confidence. Although the uneven illumination downgrades the performance of all methods, both of our methods obtain lower HTER in comparison with other methods, except the method with +120lx under iPad photo and 2D mask in the close background distance setting. It indicates that our model is robust in different ambient illuminations. One possible explanation is that the influence of the ambient illumination is reduced since the illuminance of the additional flash light is much stronger. In contrast, the EB method is the most sensitive to the ambient illumination change since the detection of eye blinking requires a clear image of the eyes.

Since EB, OFF and DLS methods only rely on the face region, their performances are independent of the distance between the subject and the background. HTER of all methods with far background distance are generally lower than the

TABLE II

AVERAGE HTER (%) OF THE PROPOSED MODELS WITH +120lx AND +160lx, AND THE COMPARISON METHODS IN DIFFERENT ENVIRONMENTAL SETTINGS (N: NORMAL AMBIENT ILLUMINANCE, U: UNEVEN AMBIENT ILLUMINANCE, C: CLOSE BACKGROUND DISTANCE, F: FAR BACKGROUND DISTANCE, AVG: AVERAGE HTER OF ALL SETTINGS)

Methods	Attack Types and Settings														AVG
	Photo(Paper)		Photo(iPad)		Video		2D Mask				Curved Mask				
	N	U	N	U	N	U	C+N	C+U	F+N	F+U	C+N	C+U	F+N	F+U	
Proposed (+120lx)	1.23	1.36	0.83	1.42	1.95	1.74	3.79	1.90	0.99	**0.68**	1.84	2.26	**0.33**	1.47	1.56
Proposed (+160lx)	**1.03**	**1.13**	**0.50**	**0.66**	1.07	1.95	3.90	**1.12**	**0.35**	0.88	**1.01**	1.62	0.70	**0.51**	**1.17**
EB^	10.21*◇	7.95*◇	9.97*◇	8.09*◇	17.56*◇	9.93*◇	12.16*◇	8.33*◇	12.16*◇	8.33*◇	15.57*◇	10.36*◇	15.57*◇	10.36*◇	11.18
OFF^	9.97*◇	6.17*◇	4.94*◇	5.15*◇	12.23*◇	11.00*◇	2.06*◇	2.74◇	2.06*◇	2.74	6.16*◇	2.83*◇	2.15*◇	2.83*◇	5.24
DLS^	2.88*◇	4.52*◇	1.85*◇	2.72*◇	3.64*◇	5.42*◇	2.78*	4.54*◇	2.78*◇	4.54*◇	5.78*◇	5.40*◇	5.78*◇	5.40*◇	4.15
TI°	1.19	1.19◇	2.47*◇	2.47*◇	3.66*◇	3.66*◇	1.19*◇	1.19*◇	1.19*◇	1.19*◇	1.22*◇	**1.22***◇	1.22*◇	1.22*◇	1.73
TI_{att}°△	1.28*◇	1.28*◇	3.66*◇	3.66*◇	6.04*◇	6.04*◇	5.23*◇	5.23*◇	5.23*◇	5.23*◇	2.25*◇	2.25*◇	2.25*◇	2.25*◇	3.71
LBP	6.33*◇	5.30*◇	2.00	3.25*	3.10*◇	3.83◇	1.40*◇	1.80*◇	1.28◇	1.36*	1.68*	3.12*◇	0.66	3.08	2.73
DS	1.80*◇	2.70*◇	1.67*◇	1.17*◇	4.10*◇	5.73*◇	5.20*◇	2.14*◇	4.13*◇	3.69*◇	1.36*◇	2.35*◇	1.20*◇	2.40*◇	2.83
LBP_F (+120lx)	3.27*◇	1.25*	1.25*◇	6.67*◇	2.53	3.87*◇	3.82*◇	5.75*◇	1.00*◇	2.50*◇	**1.01***	3.04	5.25*◇	1.75*◇	3.07
LBP_F (+160lx)	2.02*◇	1.76	6.51*◇	2.26*◇	1.75*◇	2.28*◇	2.53*◇	5.12*◇	1.25	6.05*◇	4.60	5.62	4.86*◇	2.00*◇	3.47
DS_F (+120lx)	4.04*◇	4.59*◇	4.28*◇	1.27	2.51◇	1.51	4.56*◇	3.53*◇	7.34*◇	2.51*◇	1.76*◇	5.29*◇	4.58*◇	2.78*◇	3.61
DS_F (+160lx)	6.09*◇	5.08*◇	3.04*◇	2.51*◇	1.52*	1.77	6.30*◇	3.55*◇	5.56*◇	2.79*◇	6.06*◇	4.29*◇	7.10*◇	5.29*◇	4.35
LBP+LBP_F (+120lx)	5.29*◇	3.97*◇	1.67*	2.48*◇	2.60*◇	3.64*◇	2.96◇	1.64*◇	1.25*◇	1.75*◇	1.86◇	3.02*◇	0.69◇	1.72◇	2.47
LBP+LBP_F (+160lx)	4.99*◇	3.58*◇	1.86*◇	3.27*◇	2.47*◇	3.44*◇	2.10*◇	1.97*◇	0.99*◇	1.15*	4.47*◇	3.18*◇	0.55◇	1.34◇	2.53
DS+DS_F (+120lx)	3.03*◇	2.79*◇	2.77*◇	1.51*◇	**0.75**	**1.01***	2.51◇	2.78*	2.78*◇	2.51*◇	2.01*	2.26◇	5.05*◇	2.00*◇	2.41
DS+DS_F (+160lx)	2.54*◇	1.53*◇	1.26*◇	1.51*◇	1.76	1.25*	**1.00***◇	3.02*◇	1.26*◇	2.26*◇	1.51*	2.02*◇	2.53*◇	3.80*◇	1.95

◇ Statistically significant difference with 95% confidence in comparison with our proposed method (+120lx) using the Student's t-test.
* Statistically significant difference with 95% confidence in comparison with our proposed method (+160lx) using the Student's t-test.
^ The method is independent to background distance.
° The method is independent to both background distance and environmental illuminance.
△ The temperature of the 2D spoofing attack is raised intensionally in this method.

ones with close background distance. It is because the depth information is more easily detected with the increase of the background distance. In both scenarios, the proposed models maintain stable and satisfying performance.

The significant temperature difference between a real face and the spoofing attacks causes TI to achieve a satisfying performance and the result is more accurate than other existing face liveness detection methods. However, HTER of TI is still lower than the one for our proposed methods. Moreover, if an adversary raises the temperature of the object in order to reduce the difference between a real face and the attack, HTER of TI increases dramatically. The results are shown in the row of TI_{att} in Table II. It indicates a security hole of TI which should be further studied to increase its robustness in an adversarial environment.

We further investigate whether or not the use of flash image will improve the accuracy of a face liveness detection method. HTER of LBP and DS are compared with the one of LBP and DS on flash images (LBP_F (+120lx), LBP_F (+160lx), DS_F (+120lx), and DS_F (+160lx)), combination of LBP and LBP_F with average fusion (LBP+LBP_F (+120lx), LBP+LBP_F (+160lx)) with average fusion, and combination of DS and DS_F with average fusion (DS+DS_F (+120lx), DS+DS_F (+160lx)) in Table II.

The experimental results show that the method using only flash images is not consistently better the one with non-flash images. For LBP, flash images improve the detection of photo and video attacks, *i.e.* the average HTER on photo and video attacks of LBP_F is lower than 1.46 under normal ambient illuminance. However, LBP with flash images becomes less accurate on 2D and curved mask attacks than LBP with non-flash images. In 8 out of 14 cases, LBP_F with +120lx and +160lx flash images is better than LBP. It is 7 out of 14 cases for LBP_F with +160lx flash images. However, the average HTER of LBP (2.73) is slightly lower than the one of LBP_F (3.07 for +120 and 3.47 for +160). This indicates that LBP with flash images is not robust consistently, which explains why additional structure features are considered in our proposed method. For DS, the contribution of flash images is less insignificant. Only 3 out of 14 cases and 1 out of 14 cases show that DS_F (+120lx) and DS_F (+160lx) are better than DS with 95% significant confidence. This may be because DS focuses on weak light diffusion on a human face, which becomes difficult to capture with flash.

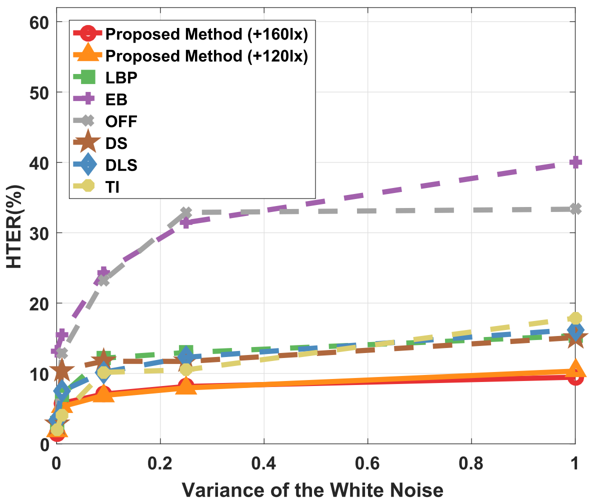


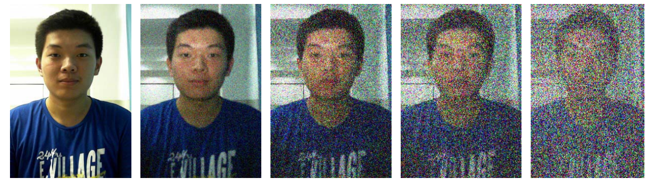
Fig. 9. Average HTER (%) of the proposed methods with +120lx and +160lx, and the comparison methods on images contaminated by the white noise with difference variances.

While considering the fusion of the methods with flash and non-flash images, HTER of LBP+LBP_F and DS+DS_F is significantly lower than the one for LBP, DS, LBP_F and DS_F generally. The results suggest the importance of considering both flash and non-flash images. The utilization of both images may provide a useful comparison to indicate whether the subject is from a spoofing attack. Although LBP+LBP_F and DS+DS_F achieves relatively good performance, their HTER is higher than the one for our methods (both with +120lx and +160lx) in all cases except video attack under normal and uneven illuminance, and 2D mask under illuminance for close background distance.

In summary, the experimental results demonstrate that the proposed method with +160lx successfully outperforms other comparison methods under various types of spoofing attacks. Although a flash with lower intensity is used, our method with +120lx still achieves satisfying results which are better than other methods generally. The performance of our method is also less sensitive to different environmental factors including the background distance and the ambient illuminance.

3) Comparison With Existing Methods With Noisy Images: The robustness of the face liveness detection to noisy images is evaluated. Only the close background distance and normal ambient illuminance are considered in this comparison. The average HTER of the detection method for all five types of attacks is calculated. All detection methods are trained with untainted training set. The white Gaussian noise with the variance = 0.01, 0.09, 0.25 and 1, and the mean = 0 are added to each testing sample, which has been normalized to the interval [0, 1]. The examples of the noisy images are shown in figure 10.

The experimental results shown in figure 9 suggest that the performances of all methods suffer from the noise, *i.e.* HTER increases with the noise. There is no significant difference between the performance of our methods with +120lx and +160lx on images with the white noise with different variances. They achieve the most robust performance among all methods. As real faces and 2D spoofing attacks are clearly



(a) (b) (c) (d) (e)

Fig. 10. Examples of noise images with different variances: (a) 0; (b) 0.01; (c) 0.09; (d) 0.25; (e) 1.

TABLE III
AVERAGE RUNNING TIME OF FEATURE EXTRACTION AND CLASSIFICATION FOR THE PROPOSED METHOD AND COMPARISON METHODS

	Our Method	LBP	EB	OFF	DS	DLS	TI
Feature Extraction(s)	**0.04**	0.17	22.93	1.71	0.16	43.13	0.07
Classification(s)	0.20	0.19	**0.01**	**0.01**	0.16	0.19	0.20
Total Time(s)	**0.24**	0.36	22.94	1.78	0.32	43.32	0.27

separated in our system, the white noise with large variance does not affect our results dramatically, *i.e.* their HTER values increase slightly with the increase of variance of the white noise. The results indicate that our model is robust to the while noise even though only flash with weak intensity is used. HTER of TI, DS, DLS and LBP increases more slowly than EB and OFF. Since EB and OFF highly depend on pixel-level analysis, they are noise sensitive. This observation agrees with the result in the previous section.

4) Computational Complexity: The computational complexity of the methods in terms of the average running time of feature extraction and the classification are given in Table III. The proposed method has the lowest computational complexity of feature extraction since only LBP_FI as well as the standard deviation and mean values are extracted. As different from the traditional LBP method which extracts the value from the whole picture, LBP_FI of our model only measures the face region which is much smaller than the original image. EB and DLS cost the most extraction time because complicated features are extracted from hundreds of frames. As the extraction of LBP and intensity histogram is required for TI, its time complexity is slightly larger than the one of the proposed method. The classification times of all methods are similar except EB and OFF since they only consider a single, one-dimensional feature. In conclusion, although two images are processed in our method, its time complexity is still relatively low in comparison with other detection methods.

5) Effectiveness of the Descriptors of the Proposed Method: The discriminant ability of the descriptors in the proposed method is evaluated in this section. A classifier is trained using one combination of features each time. The settings such as the close background distance, normal ambient illuminance and +120lx additional illuminance are considered in this experiment. From the results given in Table IV, LBP_FI is the most critical descriptor which affects the performance of

TABLE IV

AVERAGE HTER (%) OF THE PROPOSED MODEL WITH +120lx AND DIFFERENT FEATURE COMBINATIONS IN THE CLOSE BACKGROUND DISTANCE AND NORMAL AMBIENT ILLUMINANCE. DESCRIPTOR: ① LBP_FI, ② SD_FIC, ③ M_BIC AND ④ SD_BIC

Attack Type	Feature Combinations														
	①	②	③	④	①②	①③	①④	②③	②④	③④	①②③	①②④	①③④	②③④	①②③④
Photo (Paper)	2.50	20.29	9.48	23.86	1.91	2.50	1.67	11.17	16.71	10.17	1.67	2.50	2.67	10.83	**1.62**
Photo (iPad)	1.83	15.83	3.33	17.47	2.67	2.67	2.92	1.83	5.41	1.94	1.67	2.5	1.74	2.50	**0.84**
Video	1.83	11.17	2.67	12.30	1.91	1.67	1.83	1.67	6.17	2.50	1.83	1.67	1.83	1.00	**0.83**
2D Mask	4.24	17.91	26.62	24.66	6.00	4.33	6.35	15.73	7.89	18.67	6.66	5.17	5.72	6.89	**3.10**
Curved Mask	4.33	17.93	25.58	24.67	5.50	3.91	4.50	16.39	14.39	13.11	2.91	2.65	2.58	11.02	**1.91**

TABLE V

AVERAGE HTER (%) OF THE PROPOSED MODELS WITH +120lx USING LBP_FI, DS_FI AND DOG_FI IN THE CLOSE BACKGROUND DISTANCE AND NORMAL AMBIENT ILLUMINANCE DESCRIPTOR: ① LBP_FI, ② SD_FIC, ③ M_BIC AND ④ SD_BIC, ⑤ DS_FI, ⑥ DOG_FI

Attack Type	Feature Combinations		
	①+②③④	⑤+②③④	⑥+②③④
Photo (Paper)	1.23	**1.08***	2.69*
Photo (iPad)	**0.83**	0.85*	1.17
Video	1.95	2.18	**1.25***
2D Mask	3.79	**3.23***	3.28
Curved Mask	1.84	1.81*	**1.77**
Average	1.93	**1.83**	2.03

* Statistically significant difference with 95% confidence in comparison with our proposed method (+120lx) using the Student's t-test.

TABLE VI

AVERAGE HTER (%) OF OUR METHOD WITH +120lx TRAINED WITH DIFFERENT KINDS OF ATTACKS

Training \ Test	Paper Photo	iPad Photo	Video	2D Mask	Curved Mask	All
Photo (Paper)	**1.23**	2.73	2.73	7.31	8.27	10.71
Photo (iPad)	1.77	**0.83**	0.97	4.59	9.13	9.28
Video	2.59	2.63	**1.95**	5.45	8.01	8.42
2D Mask	2.68	4.45	4.45	**3.79**	5.56	5.19
Curved Mask	1.86	2.68	2.59	1.86	**1.84**	4.51
All	0.95	**0.00**	0.92	0.86	0.85	**0.00**

the proposed model significantly. HTER of classifiers with any combination containing LBP_FI is lower than 6.7%. Furthermore, M_BIC also plays an important role in detecting attack of iPad and video where HTER of the classifiers with any combination of M_BIC is lower than 3.33%. It may be because the severe reflection of an iPad screen increases the mean value of the background region, which makes these two types of attacks more differentiable from normal faces. The descriptors SD_FIC and SD_BIC perform badly individually. For instance, the HTER of using only SD_FIC and SD_BIC is larger than 11% and 12% respectively for all attacks. However, HTER of our model using all descriptors is the lowest in each row, which suggests that although an individual descriptor may not perform well, it works well with other descriptors as a group and every one of them has a positive impact on the 2D spoofing attack detection.

Our model is also evaluated using other descriptors. LBP, which plays a key role in our model, is replaced by more advanced features, *i.e.* DS [18] and DoG [14]. Similar to LBP_FI described in Sec III-A.1, DS and DoG are applied to the image with flash in our model, named DS_FI and DoG_FI. Table V shows HTER of our original model, and our revised models in which LBP_FI is replaced by DS_FI and DoG_FI.

As DS focuses on the structure difference of the subject's face, our method using DS_FI has more satisfying performance under paper photo and 2D mask attacks than our original method. However, our original model achieves lower HTER than DS_FI in other attacks. On the other hand, the models using LBP_FI are better in photo attacks but worse in video and mask attacks than the ones using DoG_FI. The difference on HTER of the models using LBP_FI, DS_FI, and DoG_FI is less than 1%, *i.e.* they have similar performance. However, by considering its short feature extraction time, LBP_FI is a suitable feature for our model.

6) Partial Knowledge on the Attack Types: The face liveness detection may be invaded by an unseen attack in reality. In this section, we assume that the defenders know a 2D spoofing attack is used but not the type. The proposed method is trained by one of the attacks and then is evaluated by another. We consider the scenario with the close background distance and normal environmental illuminance. +120lx additional illuminance is used in our model. The results are displayed in Table VI. Each row represents our method trained by one type of attack while each column is the evaluation using the test set with another type of attack. When all types of attacks are used in the training (test) phase, the row and the column are named by "All".

The performance of our method drops when the training and test set contain different types of attacks. The five 2D spoofing attacks applied in the experiment can be categorized into two types: 1) photo & video attack, and 2) mask attack. When the attacks in the training and test set are in the same category, our method maintains a good performance. However, HTER of our model is larger when the training and test set are different, except 2D mask attack. For example, for the model using a training set with paper photo attack, its HTER on the test set with iPad photo attack (2.73%) is much lower than the one with 2D mask attack (7.31%). The classifier using a training set with 2D mask attack detects paper photo attack more accurately than 2D mask attack in the test phase. This is mainly because paper photo attack is similar to 2D mask attack but easier to be identified. This observation in general agrees with other classification problems, namely, the similarity between training and test sets affects the performance of detection.

When the proposed method is trained by using all kinds of attacks, the performance of classifying each attack is satisfying, which is slightly worse than the one trained with the same attack. Moreover, the HTER value of classifying all attacks is 0.0%, which is the lowest value among all methods trained with one attack. This result demonstrates that our method can handle a complicated situation arising from several kinds of attacks. If all kinds of 2D spoofing attacks are obtained in advance, our method can protect the system effectively.

V. Conclusion and Future Work

A face liveness detection method against 2D spoofing attack using flash is proposed in this paper. The descriptors of the texture (*i.e.* LBP_FI) and structure analysis (*i.e.* SD_FIC, M_BIC and SD_BIC) are carefully designed to capture the difference from two images of the subject, one with flash and the other without flash. Our method has satisfying performance because flash enhances the differences between legitimate users and attacks. The conceptual discussion is also given based on the Lambertian reflectance law. In contrast to the existing methods, the proposed model combines the advantage of the software and hardware approaches which are high accuracy, high robustness, low computational complexity and low setup cost.

A dataset containing 50 subjects with 2D spoofing attacks, including paper photo, iPad photo, video, 2D mask and curved mask attack, are collected. In order to compare with the thermal image method, thermal images of 21 subjects with real and five types of attacks are also collected. Our method is also compared experimentally with five software-based and one hardware-based liveness detection methods. The experimental results show that the proposed method is better in terms of accuracy and running time. In addition, the robustness of our method to noisy images and different environmental settings including the background distance and ambient illuminance is better than other methods.

The tradeoff of the superiority of our method is the installation of an additional hardware, *i.e.* flash. It may limit the applications of our method, *e.g.* frontal flash is not a necessary device for a smartphone. However, different from other hardware-based methods, it may not be a serious issue since the installation cost of a flash is low in comparison with other hardware used, *e.g.* a thermal camera. Moreover, flash becomes more popular and can be found in many systems recently, *e.g.* frontal flash is more popular recently due to the popularity of the selfie.

Although the illuminance of flash in our current model is no harm to human eyes and it is also much lower than the illuminance of flash used in a camera, user comfort is a concern. A possible solution to overcome this limitation is to adjust the angle of flash on a subject. If flash is not installed at the eye level, the lighting of flash will not directly irritate human eyes and a subject will feel more comfortable. The angle of flash should be determined according to not only the detection accuracy but also installation difficulty. Other robust features may be considered in our model due to the change of lighting angle.

With the promising results obtained in this study of using flash in against 2D spoofing attack, one possible future work is to focus on exploring the performance of the proposed model on the detection of more advanced attacks, such as, the 3D spoofing attacks, for instance, rigid 3D mask and 3D face models with various expressions. The reflected light from a real face and a 3D mask is expected to be different since they have different surface reflectivity. Moreover, the texture detail of the 3D masks may also be enhanced by the flash. As a result, the additional lighting should be useful to separate legitimate users from the attacks if suitable descriptors can be identified.

References

[1] A. K. Jain, A. Ross, and S. Prabhakar, "An introduction to biometric recognition," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 14, no. 1, pp. 4–20, Jan. 2004.

[2] U. Uludag, S. Pankanti, S. Prabhakar, and A. K. Jain, "Biometric cryptosystems: Issues and challenges," *Proc. IEEE*, vol. 92, no. 6, pp. 948–960, Jun. 2004.

[3] W. Zhao, R. Chellappa, P. J. Phillips, and A. Rosenfeld, "Face recognition: A literature survey," *ACM Comput. Surv.*, vol. 35, no. 4, pp. 399–458, 2003.

[4] A. F. Abate, M. Nappi, D. Riccio, and G. Sabatino, "2D and 3D face recognition: A survey," *Pattern Recognit. Lett.*, vol. 28, no. 14, pp. 1885–1906, Oct. 2007.

[5] A. Wagner, J. Wright, A. Ganesh, Z. Zhou, H. Mobahi, and Y. Ma, "Toward a practical face recognition system: Robust alignment and illumination by sparse representation," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 34, no. 2, pp. 372–386, Feb. 2012.

[6] C.-P. Wei, C.-F. Chen, and Y.-C. F. Wang, "Robust face recognition with structurally incoherent low-rank matrix decomposition," *IEEE Trans. Image Process.*, vol. 23, no. 8, pp. 3294–3307, Aug. 2014.

[7] S. Gao, Y. Zhang, K. Jia, J. Lu, and Y. Zhang, "Single sample face recognition via learning deep supervised autoencoders," *IEEE Trans. Inf. Forensics Security*, vol. 10, no. 10, pp. 2108–2118, Oct. 2015.

[8] I. Chingovska, A. R. dos Anjos, and S. Marcel, "Biometrics evaluation under spoofing attacks," *IEEE Trans. Inf. Forensics Security*, vol. 9, no. 12, pp. 2264–2276, Dec. 2014.

[9] B. Biggio, Z. Akhtar, G. Fumera, G. L. Marcialis, and F. Roli, "Security evaluation of biometric authentication systems under real spoofing attacks," *IET Biometrics*, vol. 1, no. 1, pp. 11–24, Mar. 2012.

[10] Z. Akhtar, C. Micheloni, and G. L. Foresti, "Biometric liveness detection: Challenges and research opportunities," *IEEE Security Privacy*, vol. 13, no. 5, pp. 63–72, Sep./Oct. 2015.

[11] R. Tronci *et al.*, "Fusion of multiple clues for photo-attack detection in face recognition systems," in *Proc. IEEE Int. Joint Conf. Biometrics (IJCB)*, Washington, DC, USA, Oct. 2011, pp. 1–6.

[12] I. Chingovska *et al.*, "The 2nd competition on counter measures to 2D face spoofing attacks," in *Proc. IAPR Int. Conf. Biometrics (ICB)*, Madrid, Spain, Jun. 2013, pp. 1–6.

[13] A. Anjos and S. Marcel, "Counter-measures to photo attacks in face recognition: A public database and a baseline," in *Proc. IEEE Int. Joint Conf. Biometrics (IJCB)*, Washington, DC, USA, Oct. 2011, pp. 1–7.

[14] X. Tan, Y. Li, J. Liu, and L. Jiang, "Face liveness detection from a single image with sparse low rank bilinear discriminative model," in *Proc. 11th Eur. Conf. Comput. Vis. (ECCV)*, Heraklion, Greece, Sep. 2010, pp. 504–517.

[15] D. Wen, H. Han, and A. K. Jain, "Face spoof detection with image distortion analysis," *IEEE Trans. Inf. Forensics Security*, vol. 10, no. 4, pp. 746–761, Apr. 2015.

[16] Z. Zhang, J. Yan, S. Liu, Z. Lei, D. Yi, and S. Z. Li, "A face antispoofing database with diverse attacks," in *Proc. 5th IAPR Int. Conf. Biometrics (ICB)*, New Delhi, India, Mar./Apr. 2012, pp. 26–31.

[17] I. Chingovska, A. Anjos, and S. Marcel, "On the effectiveness of local binary patterns in face anti-spoofing," in *Proc. Int. Conf. Biometrics Special Interest Group (BIOSIG)*, Darmstadt, Germany, Sep. 2012, pp. 1–7.

[18] W. Kim, S. Suh, and J.-J. Han, "Face liveness detection from a single image via diffusion speed model," *IEEE Trans. Image Process.*, vol. 24, no. 8, pp. 2456–2465, Aug. 2015.

[19] S. Chakraborty and D. Das. (2014). "An overview of face liveness detection." [Online]. Available: https://arxiv.org/abs/1405.2227

[20] J. Li, Y. Wang, T. Tan, and A. K. Jain, "Live face detection based on the analysis of Fourier spectra," *Proc. SPIE*, vol. 5404, pp. 296–303, Apr. 2004.

[21] J. Määttä, A. Hadid, and M. Pietikäinen, "Face spoofing detection from single images using micro-texture analysis," in *Proc. Int. Joint Conf. Biometrics (IJCB)*, Washington, DC, USA, Oct. 2011, pp. 1–7.

[22] W. Bao, H. Li, N. Li, and W. Jiang, "A liveness detection method for face recognition based on optical flow field," in *Proc. Int. Conf. Image Anal. Signal Process. (IASP)*, Taizhou, China, Apr. 2009, pp. 233–236.

[23] T. Choudhury, B. Clarkson, T. Jebara, and A. Pentland, "Multimodal person recognition using unconstrained audio and video," in *Proc. Int. Conf. Audio Video-Based Person Authentication*, 1999, pp. 176–181.

[24] J.-W. Li, "Eye blink detection based on multiple Gabor response waves," in *Proc. Int. Conf. Mach. Learn. Cybern. (ICMLC)*, vol. 5. Jul. 2008, pp. 2852–2856.

[25] H. Yu, T.-T. Ng, and Q. Sun, "Recaptured photo detection using specularity distribution," in *Proc. 15th IEEE Int. Conf. Image Process. (ICIP)*, Oct. 2008, pp. 3140–3143.

[26] J. Galbally, S. Marcel, and J. Fierrez, "Image quality assessment for fake biometric detection: Application to iris, fingerprint, and face recognition," *IEEE Trans. Image Process.*, vol. 23, no. 2, pp. 710–724, Feb. 2014.

[27] G. Kim, S. Eum, J. K. Suhr, D. I. Kim, K. R. Park, and J. Kim, "Face liveness detection based on texture and frequency analyses," in *Proc. 5th IAPR Int. Conf. Biometrics (ICB)*, New Delhi, India, Mar./Apr. 2012, pp. 67–72.

[28] S. Tirunagari, N. Poh, D. Windridge, A. Iorliam, N. Suki, and A. Ho, "Detection of face spoofing using visual dynamics," *IEEE Trans. Inf. Forensics Security*, vol. 10, no. 4, pp. 762–777, Apr. 2015.

[29] G. Chetty, "Biometric liveness detection based on cross modal fusion," in *Proc. 12th Int. Conf. Inf. Fusion (FUSION)*. Seattle, WA, USA, Jul. 2009, pp. 2255–2262.

[30] K. Kollreider, H. Fronthaler, and J. Bigun, "Evaluating liveness by face images and the structure tensor," in *Proc. 4th IEEE Workshop Automat. Identificat. Adv. Technol.*, Buffalo, NY, USA, Oct. 2005, pp. 75–80.

[31] A. Pinto, W. R. Schwartz, H. Pedrini, and A. Rocha, "Using visual rhythms for detecting video-based facial spoof attacks," *IEEE Trans. Inf. Forensics Security*, vol. 10, no. 5, pp. 1025–1038, May 2015.

[32] Z. Zhang, D. Yi, Z. Lei, and S. Z. Li, "Face liveness detection by learning multispectral reflectance distributions," in *Proc. FG*, Santa Barbara, CA, USA, Mar. 2011, pp. 436–441.

[33] W. Liu, "Face liveness detection using analysis of Fourier spectra based on hair," in *Proc. Int. Conf. Wavelet Anal. Pattern Recognit. (ICWAPR)*, Lanzhou, China, Jul. 2014, pp. 75–80.

[34] T. Ojala, M. Pietikäinen, and T. Mäenpää, "Multiresolution gray-scale and rotation invariant texture classification with local binary patterns," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 24, no. 7, pp. 971–987, Jul. 2002.

[35] Z. Boulkenafet, J. Komulainen, and A. Hadid, "Face spoofing detection using colour texture analysis," *IEEE Trans. Inf. Forensics Security*, vol. 11, no. 8, pp. 1818–1830, Aug. 2016.

[36] M. Smiatacz, "Liveness measurements using optical flow for biometric person authentication," *Metrol. Meas. Syst.*, vol. 19, no. 2, pp. 257–268, 2012.

[37] A. Anjos, M. M. Chakka, and S. Marcel, "Motion-based countermeasures to photo attacks in face recognition," *IET Biometrics*, vol. 3, no. 3, pp. 147–158, Sep. 2014.

[38] H. K. Jee, S. U. Jung, and J. H. Yoo, "Liveness detection for embedded face recognition system," in *Proc. World Acad. Sci., Eng. Technol.*, Vienna, Austria, Dec. 2006, pp. 29–32.

[39] G. Pan, L. Sun, Z. Wu, and S. Lao, "Eyeblink-based anti-spoofing in face recognition from a generic webcamera," in *Proc. IEEE 11th Int. Conf. Comput. Vis. (ICCV)*, Rio de Janeiro, Brazil, Oct. 2007, pp. 1–8.

[40] G. Chetty and M. Wagner, "Multi-level liveness verification for face-voice biometric authentication," in *Proc. Biometrics Symp., Special Session Res. Biometric Consortium Conf.*, Baltimore, MD, USA, Sep. 2006, pp. 1–6.

[41] K. Patel, H. Han, and A. K. Jain, "Secure face unlock: Spoof detection on smartphones," *IEEE Trans. Inf. Forensics Security*, vol. 11, no. 10, pp. 2268–2283, Oct. 2016.

[42] T. I. Dhamecha, A. Nigam, R. Singh, and M. Vatsa, "Disguise detection and face recognition in visible and thermal spectrums," in *Proc. Int. Conf. Biometrics*, Jun. 2013, pp. 1–8.

[43] A. Lagorio, M. Tistarelli, M. Cadoni, C. Fookes, and S. Sridharan, "Liveness detection based on 3D face shape analysis," in *Proc. 1st Int. Workshop Biometrics Forensics (IWBF)*, Lisbon, Portugal, Apr. 2013, pp. 1–4.

[44] S. Kim, Y. Ban, and S. Lee, "Face liveness detection using a light field camera," *Sensors*, vol. 14, no. 12, pp. 22471–22499, Jan. 2014.

[45] G. Pan, L. Sun, Z. Wu, and Y. Wang, "Monocular camera-based face liveness detection by combining eyeblink and scene context," *Telecommun. Syst.*, vol. 47, pp. 215–225, Aug. 2011.

[46] Y. Kim, J.-H. Yoo, and K. Choi, "A motion and similarity-based fake detection method for biometric face recognition systems," *IEEE Trans. Consum. Electron.*, vol. 57, no. 2, pp. 756–762, May 2011.

[47] D. A. Socolinsky, A. Selinger, and J. D. Neuheisel, "Face recognition with visible and thermal infrared imagery," *Comput. Vis. Image Understand.*, vol. 91, nos. 1–2, pp. 72–114, Jun./Aug. 2003.

[48] G. Chetty and M. Wagner, "Biometric person authentication with liveness detection based on audio-visual fusion," *Int. J. Biometrics*, vol. 1, no. 4, pp. 463–478, 2009.

[49] G. Chetty, "Biometric liveness checking using multimodal fuzzy fusion," in *Proc. IEEE Int. Conf. Fuzzy Syst.*, Barcelona, Spain, Jul. 2010, pp. 1–8.

[50] T. Wang, J. Yang, Z. Lei, S. Liao, and S. Z. Li, "Face liveness detection using 3D structure recovered from a single camera," in *Proc. ICB*, Madrid, Spain, Jun. 2013, pp. 1–6.

[51] M. Nilsson, J. Nordberg, and I. Claesson, "Face detection using local smqt features and split up snow classifier," in *Proc. IEEE Int. Conf. Acoust., Speech, Signal Process. (ICASSP)*, vol. 2. Apr. 2007, pp. II-589–II-592.

[52] T. Ahonen, A. Hadid, and M. Pietikäinen, "Face recognition with local binary patterns," in *Proc. 8th Eur. Conf. Comput. Vis.*, Prague, Czech Republic, May 2004, pp. 469–481.

[53] R. Basri and D. W. Jacobs, "Lambertian reflectance and linear subspaces," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 25, no. 2, pp. 218–233, Feb. 2003.

[54] P. Liu, F. Zafar, and A. Badano, "The effect of ambient illumination on handheld display image quality," *J. Digit. Imag.*, vol. 27, no. 1, pp. 12–18, Feb. 2014.

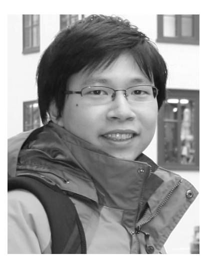
[55] *Microsoft LifeCam Studio*. [Online]. Available: http://www.microsoftstore.com/store/msusa/en_US/pdp/productID.258411900

[56] *F60M External Flash for Multi-Interface Shoe—HVL-F60M—Sony us*. [Online]. Available: https://www.sony.com/electronics/interchangeable-lens-cameras-flashes-lights/hvl-f60m

[57] *F43M External Flash for Multi-Interface Shoe—HVL-F43M—Sony us*. [Online]. Available: https://www.sony.com/electronics/ interchangeable-lens-cameras-flashes-lights/hvl-f43m

[58] *CompactXR—Seek Thermal*. [Online]. Available: http://www.thermal.com/products/compactxr/

[59] C.-C. Chang and C.-J. Lin, "LIBSVM: A library for support vector machines," *ACM Trans. Intell. Syst. Technol.*, vol. 2, no. 3, pp. 27:1–27:27, 2011.



Patrick P. K. Chan received the Ph.D. degree from The Hong Kong Polytechnic University in 2009. He is currently an Associate Professor with the School of Computer Science and Engineering, and the person in charge of machine learning and the Cybernetics Research Laboratory, South China University of Technology, Guangzhou, China. He is also a part-time Lecturer with the Hyogo College of Medicine, Japan. His current research interests include pattern recognition, multiple classifier system, biometric, computer security, deep learning, and reinforcement learning. He was a member of the governing boards of the IEEE SMC Society from 2014 to 2016. He serves as an Organizing Committee Chair of several international conferences. He was also the Chairman of the IEEE SMCS Hong Kong Chapter 14–15. He is the Counselor of the IEEE Student Branch, South China University of Technology. He is an associate editor for international journals, including *Information Sciences* and the *International Journal of Machine Learning and Cybernetics*.

Weiwen Liu received the B.S. degree in computer science and technology from the South China University of Technology in 2013. She is currently pursuing the Ph.D. degree in computer science and engineering with The Chinese University of Hong Kong. Her research interests include adversarial learning, machine learning, and machine learning algorithms.

Danni Chen received the B.S. degree from the School of Computer Science and Engineering, South China University of Technology, China, in 2016, where she is currently pursuing the M.S. degree. Her current research interests include computer vision and machine learning.

Daniel S. Yeung (F'04) received the Ph.D. degree in applied mathematics from Case Western Reserve University. He was an Assistant Professor of mathematics and computer science with the Rochester Institute of Technology, USA, as a Research Scientist with the General Electric Corporate Research Center, USA, and as a System Integration Engineer with TRW, USA. He was a Visiting Professor with the School of Computer Science and Engineering, South China University of Technology, Guangzhou, China, from 2008 to 2015. His current research interests include neural-network sensitivity analysis, data mining, and big data analytic. He was the Chairman of the Department of Computing, The Hong Kong Polytechnic University, Hong Kong, and a Chair Professor from 1999 to 2006. He is a Past President of the IEEE Systems and the Man and Cybernetics Society. He is a Co-Editor-in-Chief of the Springer *International Journal on Machine Learning and Cybernetics*.

Fei Zhang received the Ph.D. degree from the South China University of Technology, Guangzhou, China. She is currently a Lecturer with the College of Computer and Information Engineering, Henan Normal University, Xinxiang, China. Her current research interests include machine learning, computer security, and recommender system.

Xizhao Wang (M'03–SM'04–F'12) received the Ph.D. degree in computer science from the Harbin Institute of Technology in 1998. From 1998 to 2001, he was with the Department of Computing, The Hong Kong Polytechnic University, as a Research Fellow. From 2001 to 2014, he was with Hebei University as a Professor and the Dean of the School of Mathematics and Computer Sciences. He was the Founding Director of the Key Laboratory on Machine Learning and Computational Intelligence, Hebei. He was a Distinguished Lecturer of the IEEE SMCS. Since 2014, he has been a Professor with the Big Data Institute, Shenzhen University. He has edited over ten special issues and authored or co-authored over three monographs, two textbooks, and over 200 peer-reviewed research papers. As a Principle Investigator (PI) or co-PI, he has completed over 30 research projects. His research interests include uncertainty modeling and machine learning for big data. He is the previous BoG Member of the IEEE SMC Society. He was a recipient of the IEEE SMCS Outstanding Contribution Award in 2004 and the IEEE SMCS Best Associate Editor Award in 2006. He is the Chair of the IEEE SMC Technical Committee on Computational Intelligence and the General Co-Chair of the 2002–2017 International Conferences on Machine Learning and Cybernetics, co-sponsored by the IEEE SMCS. He is the Chief Editor of the *Machine Learning and Cybernetics Journal* and an associate editor of a couple of journals in related areas. He has supervised over 100 M.Phil. and Ph.D. students. According to Google scholar, the total number of citations is over 5000 and the maximum number of citation for a single paper is over 200. He is on the list of Elsevier 2015/2016 most cited Chinese authors.

Chien-Chang Hsu (M'07) received the M.S. and Ph.D. degrees from the National Taiwan University of Science and Technology in 1992 and 2000, respectively. He is currently a Professor with the Department of Computer Science and Information Engineering, Fu Jen Catholic University, Taiwan. He is also the Director of the Information Technology Center. His research interests include machine learning, intelligent systems, medical image processing, and medical informatics. He is the Chair of the Medical Informatics and Innovative Applications Program, Fu Jen Catholic University.

AUTHOR QUERIES

AUTHOR PLEASE ANSWER ALL QUERIES

PLEASE NOTE: We cannot accept new source files as corrections for your paper. If possible, please annotate the PDF proof we have sent you with your corrections and upload it via the Author Gateway. Alternatively, you may send us your corrections in list format. You may also upload revised graphics via the Author Gateway.

AQ:1 = Please provide the postal code for " South China University of Technology, Henan Normal University, Shenzhen University, and Fu Jen Catholic University."
AQ:2 = Please provide the current affiliation for "Daniel S. Yeung."
AQ:3 = Please confirm the volume no. for ref. [20].
AQ:4 = Please confirm the title and also provide the accessed date for refs. [55]–[58].

Face Liveness Detection Using a Flash Against 2D Spoofing Attack

Patrick P. K. Chan, *Member, IEEE*, Weiwen Liu, Danni Chen, Daniel S. Yeung, *Fellow, IEEE*, Fei Zhang, Xizhao Wang, *Fellow, IEEE*, and Chien-Chang Hsu, *Member, IEEE*

***Abstract*— Face recognition technique has been widely applied to personal identification systems due to its satisfying performance. However, its security may be a crucial issue, since many studies have shown that face recognition systems may be vulnerable in an adversarial environment, in which an adversary can camouflage as a legitimate user in order to mislead the system. Although face liveness detection methods have been proposed to distinguish real and fake faces, they are either time-consuming, costly, or sensitive to noise and illumination. This paper proposes a face liveness detection method with flash against 2D spoofing attack. Flash not only can enhance the differentiation between legitimate and illegitimate users, but it also reduces the influence of environmental factors. Two images are taken from a subject, one with flash and another without flash. Four texture and 2D structure descriptors with low computational complexity are used to capture information of the two images in our model. Advantages of our method include low installation cost of flash and no user cooperation required. A data set of 50 subjects collected under different scenarios is used in the experiments to evaluate the proposed method. The experimental results indicate that the proposed model performs better than existing liveness detection methods in different environmental scenarios. This paper confirms that the use of flash successfully improves face liveness detection in terms of accuracy, robustness, and running time.**

***Index Terms*— Face liveness detection, 2D spoofing attack, flash light, adversarial learning.**

I. INTRODUCTION

BIOMETRIC technology has been used widely in personal identification applications. As compared with the traditional security methods like passcodes, biometric technology brings about convenience which uses human intrinsic characteristics for individual identification [1], [2]. Face recognition is one of the most common biometric features because information from the face can be extracted easily without any physical contact. It has been successfully demonstrated in many personal identification applications, *e.g.* law enforcement, surveillance, information security, smart card authentication and entertainment [3]–[7].

Since traditional face recognition systems do not consider the existence of an adversary, many studies have revealed that these systems are vulnerable to spoofing attacks [8]–[10] in which an attacker obtains an illegitimate access to a system by camouflaging as an authorized person. A well-known example is a *2D spoofing attack*, which misleads a system by using a 2D facial duplicate of a valid user. As an image or a video of a person is easily obtainable and highly reproducible [11], [12], 2D spoofing attack is one of the most common attacks. There are three types of 2D spoofing attacks, namely photo attack, video attack and mimic mask attack. Photo attack evades the detection by using a picture of a legitimate user on a piece of paper [13], [14], or an electronic screen [15], while video attack misleads the system by using a video of an authorized person on electronic devices [16], [17]. In mimic mask attack, an adversary camouflages as an authorized person by wearing a 2D mask [18].

Face liveness detection [19], which is also referred to *face spoofing detection*, has been devised to defend against 2D spoofing attack. Face liveness detection determines whether an image is taken from a real or fake subject before face recognition process starts. Suspected images are filtered and will not be passed to the recognition system.

Previous works on face liveness detection mainly focus on *software-based methods* which analyze liveness clues, including texture [20], [21], structure information [22], [23] and liveness sign [24], of the subjects, and quality of captured images [15], [25], [26]. These methods are generally sensitive to environmental factors [19], [27], for instance, bad illumination condition and noisy images. Thus, their detection accuracy decreases significantly under such circumstances. In addition, computational complexity of calculating some liveness clue is high, *e.g.* facial dynamic is calculated based on consecutive frames [28]. Although asking users to speak [29] or shake their heads [30] improves the accuracy of the detection, it also reduces efficiency due to longer detection duration and uncooperative users. On the other hand, a device is embedded in a recognition system in *hardware-based methods* [31], [32] to capture additional information of the

Manuscript received November 23, 2016; revised March 16, 2017, June 27, 2017, and August 18, 2017; accepted September 17, 2017. This work was supported in part by Fundamental Research Funds for Central Universities under Grant 2015ZZ092 and in part by the National Training Program of Innovation and Entrepreneurship of China under Grant 201510561067. The associate editor coordinating the review of this manuscript and approving it for publication was Prof. Domingo Mery. *(Corresponding author: Fei Zhang.)*

P. P. K. Chan and D. Chen are with the School of Computer Science
AQ:1 and Engineering, South China University of Technology, Guangzhou, China (e-mail: patrickchan@ieee.org; conniechen9469@gmail.com).

W. Liu is with the Department of Computer Science and Engineering, The Chinese University of Hong Kong, Hong Kong (e-mail: patrickchan@ieee.org).

AQ:2 D. S. Yeung is with ???

F. Zhang is with the College of Computer and Information Engineering, Henan Normal University, Xinxiang, China (e-mail: zhangfei@htu.edu.cn).

X. Wang is with the College of Computer Science and Software Engineering, Shenzhen University, Shenzhen, China (e-mail: xizhaowang@ieee.org).

C.-C. Hsu is with the Computer Science and Information Engineering, Fu Jen Catholic University, Taipei, Taiwan (e-mail: cch@csie.fju.edu.tw).

Color versions of one or more of the figures in this paper are available online at http://ieeexplore.ieee.org.

Digital Object Identifier 10.1109/TIFS.2017.2758748

TABLE I
SUMMARY OF EXISTING METHODS AGAINST 2D SPOOFING ATTACK

Category	Sub-category	Description	Typical Algorithms	Pros	Cons
Software-based	Texture	Capture difference on visual and tactile quality between real and fake faces	local binary patterns(LBP) [34], Fourier analysis [20], color texture analysis [35], etc	Low implementation cost and low time complexity	Easily affected by illumination condition, noise and image quality
	Structure Information	Capture difference of structure properties between 3D real faces and 2D-planar attack	diffusion speed [18], facial feature trajectories [23], defocusing techniques [18], optical flow [22], [36], [37], etc	Relatively high detection accuracy	High time complexity, sensitive to illumination and image quality
	Liveness Sign	Capture natural human movements	Detection of eye blinking [24], [38], [39], head rotation [30] and lip movements [40]	Performs well in attacks with no human dynamics, like photo attack and mask attack	Fail to evade video attack, long detection time, high space and time complexity
	Image Quality Analysis	Analyze the quality of the real face and 2D spoof face images	Analysis of image specularity distribution [25], image distortion [15], [41] and general features [26]	Good generalization ability to various scenarios	Device dependent; Attack media with high resolution may fool the detection system
	Hybrid Methods	Combine different kinds of information to assist the detection	DMD-LBP-SVM, which combines texture and structure information [28]	Substantial information makes the detection more accurate	Longer time for feature processing leads to low detection efficiency
Hardware-based		Use additional hardware to measure the properties of a live face, like temperature and the reflectance of the subject	Infared camera [42], 3D camera, multiple 2D cameras [43] , light field camera [44], etc	High detection accuracy	High setup and maintenance cost

subjects, *e.g.* temperature. Nevertheless, some of the additional hardware is costly and difficult to install. Our preliminary study [33], which only analyzes the difference of the hair on foreheads between real and fake faces, showed that flash increases the differentiation between a legitimate person and the 2D spoofing attack. However, the study only focused on video attack in a particular environmental setting in which the ambient illumination is normal, and the distance between the camera and the background is short. The usefulness of flash on detecting other 2D spoofing attacks remains unclear. Moreover, the proposed model is sensitive to the hair on the forehead and may not be practical since users have different hair styles. Therefore in this paper we provide a complete investigation on how the use of flash can improve 2D spoofing attack detection. The literature review of face liveness detection and also 2D spoofing attack is introduced in Section II.

In Section III, a model of face liveness detection using flash to defend against photo, video and also mimic mask attacks will be elaborated. In the proposed model, a pair of images is taken from a subject in the detection, one with flash and the other without flash. Features of our method are carefully designed in order to provide accurate and robust prediction with low time complexity. The descriptor based on uniform local binary patterns is applied to measure the textural information from the face, and another three descriptors are proposed to capture the structure information of a face using the standard deviation and the mean of grayscale difference between the images with and without flash.

Then, the subject is classified as either legitimate or malicious class based on the difference between the images with and without flash measured by the four descriptors. Unlike hardware-based methods, our method requires only flash which is economical and easy to install in existing face recognition systems. The proposed method is expected to be more accurate and robust than the software-based method since flash enhances the differentiation between real and fake faces and reduces the influence of ambient illumination. In addition, the time complexity of extracting the four descriptors is low and no user cooperation is required. Our method takes advantage of both software and hardware based methods. The discussion on the reasons why considering the difference between the images with and without flash is helpful in face liveness detection based on the Lambertian reflectance law is also provided.

In Section IV, the performance of the proposed model is then evaluated and compared with other well-known face liveness detection methods under different environmental settings, including background distance and ambient illumination. The procedure of the dataset collection is also described. Finally, the conclusion and future work are given in Section V.

II. LITERATURE REVIEW

Existing face liveness detection methods against the 2D spoofing attack are briefly introduced in this section. According to the requirement of an additional device, face liveness detection methods can be categorized into software-based and hardware-based method respectively. The pros and cons in accuracy, time complexity, implementation cost and convenience to users will also be discussed. Table I summarizes the existing 2D spoofing attack detection methods.

Software-based method is the most widely used face liveness detection method. It determines whether a target is of the real face based on the information of the captured images, that is, the texture, structure information, liveness sign and image quality, without using additional hardware device. The light reflection of real human skin is different from the one displayed on a 2D-planar object, *i.e.* a paper or a mobile, in 2D spoofing attack. This difference in the visual and tactile quality is captured by *texture-based methods*. The well-known example is local binary patterns (LBP) [34] which labels the pixels of an image by thresholding the neighborhood of each pixel to represent the local texture information with the property of invariance to monotonic grayscale transformation. Generally, an image can be divided into several blocks, and LBP histograms are extracted individually. For each block, the LBP code of a pixel (x_c, y_c) is calculated using bilinearly interpolating values at non-integer sampling points in its neighborhood, as shown in (1).

$$LBP_{P,R}(x_c, y_c) = \sum_{i=0}^{P-1} g(p_i - p_c) \times 2^i, \tag{1}$$

where p_c is the gray value of the pixel (x_c, y_c) and p_i refers to the gray value of the i^{th} pixel. P and R are parameters of LBP, which represent P sampling points on a clockwise circle of radius R for each pixel's neighborhood. The function $g(z)$ is a threshold function, which outputs 1 when z is non-negative; otherwise, outputs 0. The occurrences of LBP codes are represented by a histogram. The numbers of occurrence are applied as input vectors for training.

The advanced LBP feature, referred to uniform LBP feature [34] ($LBP^{u2}_{P,R}$), is also proposed to reduce the dimensionality of the original LBP feature, which has been widely adopted in face liveness detection recently. An LBP code is uniform if it contains at most two bitwise transitions from 0 to 1 or vice versa. Each uniform LBP code is considered individually, and the rest of the non-uniform ones are grouped into one bin in the histogram. As a result, time complexity is significantly reduced since the non-uniform LBP codes are ignored. Another example of texture-based methods is the color texture of analyzing both luminance and chrominance channels which also exhibit effectiveness in 2D spoofing detection [35]. Difference of Gaussians (DoG) [14], which is a bandpass filter considering two Gaussian functions with different variances, has also been applied to improve the accuracy of the face liveness detection by removing the variant lighting in a face image. Fourier analysis [20] measures the frequency domain of face images, which is another texture information. The major drawback of a texture-based method is that its performance is highly affected by illumination condition and the quality of the input image [27]. Although the implementation cost and the time complexity are relatively low, some unexpected factors like uneven illumination and camera noise can degrade the performance significantly.

Structure information, which reveals information of the 3D structure of a subject from the projected 2D image, is also used in some detection methods. Illumination of 2D surface diffuses more slowly than that of 3D since its intensity is more evenly distributed. Diffusion is measured by the features of local speed patterns for the Diffusion Speed method (DS) [18] in order to detect a live face. Thus it is faster due to non-uniformity of the 3D surface. In addition, the depth of a face is analyzed by the facial feature trajectories [23] and the defocusing technique [18], which is a common technique for structure information. Several works on different movement patterns of 2D planes and 3D objects by optical flow fields are also captured [22], [36], [37]. The major drawbacks of these methods are high time complexity, sensitivity to the illumination and the quality of the images [36].

Some studies which focus on *liveness sign*, usually refer to the natural human movements. For example, eye blinking [24], [38], [39], head rotation [30] and lip movement [40] are common ones. Obviously, methods of this kind are designed specifically for image attacks. However, video attack is able to evade these methods easily [45], [46]. Moreover, a video has to be stored in order to detect a particular movement. This kind of method usually requires a longer detection time, and also larger space and computational complexity.

The quality of a face image in a 2D spooking attack may degrade since the face image is obtained by recapturing from photos and videos. *Image quality* has been used as an indicator in face liveness detection. For instance, the difference of specularity spatial distribution between a recaptured image and its original image [25], the distortion of a spoof attack image with respect to specular reflection, blurriness, chromatic moment, and color diversity [41], and the image quality based on 25 metrics [26] are studied. High Definition (HD) camera and display increase the resolution of mimic, which may increase the difficulty of detection by image quality analysis.

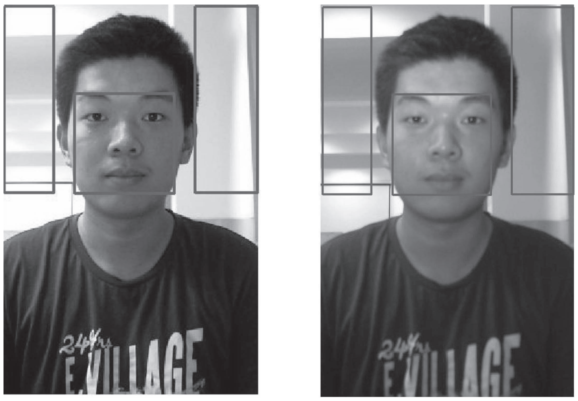
Some methods are also proposed by using different kinds of features in order to achieve higher accuracy. For instance, the features of liveness sign and texture of sequential image frames are used in dynamic mode decomposition (DMD) [28]. The model applies eye blinking, lip motion, facial expression change as well as LBP features to distinguish legitimate users from 2D spoofing attack. Another example is to apply eye blinking and background context texture to detect spoofing attack [45]. Although the time complexity is higher, the detection is usually more accurate.

In contrast, *hardware-based methods* require extra hardware to measure the additional information of subjects other than the camera of the face recognition system. A thermal camera, which has been successfully applied to face recognition [47], captures temperature and reflectance distribution of a subject. The Intensity and Texture Encoder (ITE) features [42] containing LBP and intensity histogram to detect non-biometric patches are extracted from a thermal image; a 3D camera or multiple 2D cameras [43] can be used to generate the 3D model of the subject; and a light field camera captures the light distribution of the subject [44]. Although hardware-based methods usually outperform software-based methods, the setup cost of extra devices is also much higher [1], [3].

Some detection methods need the cooperation of users. The users have to complete certain tasks during the detection process. For example, the user is required to speak for the audio-visual matching process [29], [48], [49], and to rotate the head for the 3D structure recovering process [50]. These methods achieve more accurate results at the cost of user inconvenience. However, the detection time needed is normally longer than that without user cooperation requirement.

III. Liveness Detection Method Based on Flash and No Flash Image Pairs

The proposed liveness detection method which takes advantages of both software and hardware based methods is



(a) (b)

Fig. 1. Examples of result of the face and the background extraction. The center rectangle and the rectangles on both sides of each image are the face and the background region: (a) Non-flash image; (b) Flash image.

Algorithm 1 Procedure of Feature Extraction of the Proposed Model

Input: I_n: the non-flash image; I_f: the flash image

Output: LBP_FI, SD_FIC, M_BIC and SD_BIC descriptors

1: identify I_n^F and I_n^{BG} from I_n based on a face identification method;
2: identify I_f^F and I_f^{BG} according to the locations of I_n^F and I_n^{BG} respectively;
3: extract descriptor LBP_FI from I_n^F;
4: $D^F = I_f^F - I_n^F$;
5: descriptor SD_FIC = std(D^F);
6: calculate $D^{BG} = I_f^{BG} - I_n^{BG}$;
7: descriptor M_BIC = mean(D^{BG});
8: descriptor SD_BIC = std(D^{BG}).

introduced in this section. An additional device, flash, is applied to enhance the performance of the software based method which considers the texture analysis and the structure information. The underlying principle is to magnify the differences between real face and fake face displayed in 2D media by using flash.

During the detection, two images with and without flash, denoted as I_f and I_n, are taken for the subject. We identify the rectangle regions for the face and the background defined by the pixels in the upper right corner and in the lower left corner of the region in I_n. The face region I_n^F is firstly determined. We apply the split up Sparse Network of Winnows (SNoW) classifier [51], one of the efficient face identification methods based on Successive Mean Quantization Transform. Two background regions, denoted as I_n^{BG}, are therefore located based on the face region. Specifically, the upper right corner and the lower left corner of the rectangle region of the right I_n^{BG} are defined by the upper right corner of I_n and 20 pixels to the right of the right corner of I_n^F to avoid the hair of a subject being selected. The left I_n^{BG} is defined similarly. Finally, I_f^F and I_f^{BG} are extracted from I_f according to the locations of I_n^F and I_n^{BG} respectively. Examples of the result of the face and background extraction are shown in figure 1.

Four carefully designed descriptors including LBP_FI, SD_FIC, M_BIC and SD_BIC are extracted from both regions of the face and the background. These descriptors should be able to distinguish legitimate users and the common 2D spoofing attack efficiently, accurately and robustly. The photo attack printed on a paper, the photo attack displayed on iPad, the video attack, the 2D mask attack and the curved mask attack are considered. The curved mask attack is considered as an extension of a 2D attack since it misleads the recognition system by holding the 2D mask curly. It is more difficult to detect the curved mask attack than the 2D mask attack since the curved mask covers the face more tightly than the 2D mask attack. The descriptors are input as features to a classifier for detection. The procedure for feature extraction of the proposed model is described in Algorithm 1. A real face can be distinguished from a fake one by a classifier using the extracted features. Support Vector Machine (SVM) is used in our model due to its simplicity and satisfying performance in a two-class classification problem.

In this section, the four descriptors are firstly introduced in Section III-A. Then, the underlying rationale of the proposed model is discussed in Section III-B.

A. Descriptors of the Model

1) Uniform Local Binary Patterns on the Flash Image (LBP_FI) Descriptor: LBP analysis is applied to capture the local texture information of the face region of the image with the flash (I_f^F). The reason of using I_f^F only is that the flash increases the detail of the real face but not the fake one due to the difference between 3D and 2D surfaces. As a result, a legitimate user can be distinguished from the camouflaged one.

I_f^F is firstly separated into nine non-overlapping blocks to obtain the texture information from different regions of the image [21]. The LBP code of the pixel (x, y) in each block is then calculated. In our model, the circle of radius is set to 1 and all neighbor pixels are considered, *i.e.* $P = 8$ and $R = 1$.

Since it has been shown that the uniform LBPs account for a bit less than 90% of all patterns in this setting [52], (1) of the LBP code can be simplified as (2).

$$LBP(x_c, y_c) = \sum_{i=0}^{7} g(p_i - p_c) \times 2^i. \tag{2}$$

There are totally 59 bins including 58 uniform patterns and the one containing the rest of the non-uniform patterns. The histogram $\mathbf{H}_i$ is generated according to $LBP(x_c, y_c)$ for the i^{th} block, where $\mathbf{H}_i = (h_1, h_2, \ldots, h_{59})$ and h_j is the occurrence of a pattern in j^{th} bin. Subsequently, there are a total of 531 (*i.e.* 9×59) values in LBP_FI, as shown in (3).

$$\text{LBP\_FI} = (\mathbf{H}_1, \mathbf{H}_2, \cdots, \mathbf{H}_9) = (h_1, h_2, \cdots, h_{531}). \tag{3}$$

2) Standard Deviation of Face Intensity Change (SD_FIC) Descriptor: SD_FIC measures the grayscale intensity change of the face region caused by flash. The reflection of flash varies in the real face due to its structure information, *i.e.* the

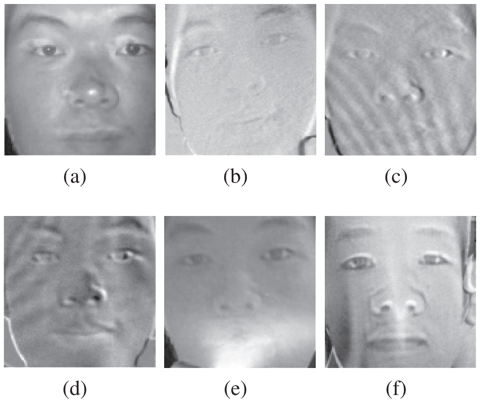


Fig. 2. Examples of the face difference images for real face and different types of attacks: (a) Real face: SD_FIC=39.45; (b) Paper photo attack: SD_FIC=19.42; (c) iPad photo attack: SD_FIC=18.52; (d) Video attack: SD_FIC=17.03; (e) 2D mask attack: SD_FIC=30.44; (f) Curved mask attack: SD_FIC=33.80.

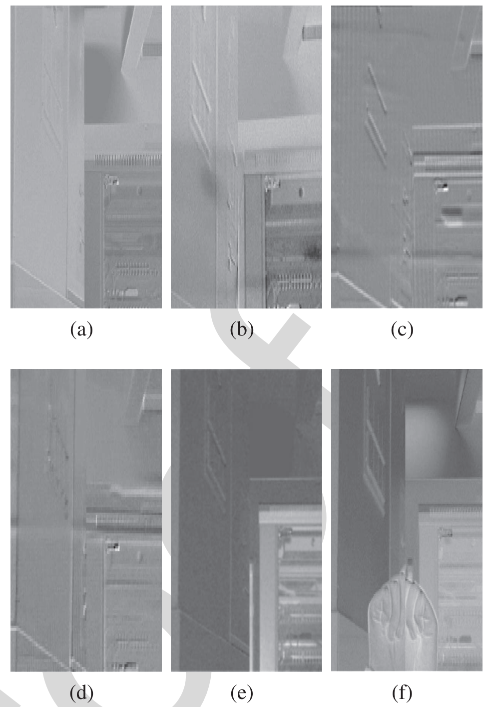


Fig. 3. Examples of the background difference images for real face and different types of attacks: (a) Real face: M_BIC=36.88, SD_BIC=24.02; (b) Paper photo attack: M_BIC=62.12, SD_BIC=25.81; (c) iPad photo attack: M_BIC=58.87, SD_BIC=17.13; (d) Video attack: M_BIC=63.24, SD_BIC=13.11; (e) 2D mask attack: M_BIC=35.57, SD_BIC=37.76; (f) Curved mask attack: M_BIC=43.88, SD_BIC=33.88.

distances between the flash and each part of the face may be different. In contrast, the reflected light of a 2D spoofing attack is more uniform. As a result, the deviation of the intensity of the real person is larger than that of a 2D spoofing attack. The standard deviation is applied to capture the change of the grayscale intensity in our model, and SD_FIC is defined as in (4).

$$\mathrm{SD\_FIC} = \sigma_{D^F} = \sqrt{\frac{\sum_{i=1}^{N}(D^F(x_i, y_i) - \mu_{D^F})^2}{N-1}}, \quad (4)$$

where μ_{D^F} and σ_{D^F} denote the mean and the standard deviation of $D^F(x, y)$ respectively, N is the number of pixels in the region and $D^F(x, y) = I_f^F(x, y) - I_n^F(x, y)$. The reason for deducting the intensity of the image without the flash light in $D^F(x, y)$ is to reduce the influence to the ambient illumination. The examples of D^F of the real face and the different types of attacks, as well as their SD_FIC values, are shown in figure 2. As discussed, the value of SD_FIC of the real face is the largest among all cases due to the intensity change on the 3D object. The paper photo, 2D mask and curved mask attacks have a larger SD_FIC than other types of attacks because a bright strip occurs in the face region.

3) Mean of Background Intensity Change (M_BIC) Descriptor: The actual background has been blocked in the photo and video attacks. As the captured background on the display media is much closer to the camera than the actual one, higher intensity of light will be reflected. We propose the M_BIC to capture this information, defined as follows:

$$\mathrm{M\_BIC} = \mu_{D^{BG}} = \frac{\sum_{i=1}^{N} D^{BG}(x_i, y_i)}{N}, \quad (5)$$

where $D^{BG}(x, y) = I_f^{BG}(x, y) - I_n^{BG}(x, y)$, $-255 \leq D^{BG} \leq 255$ and $D^{BG} \in Z$. Examples of D^{BG} of the real face and the different types of attacks are illustrated in figure 3. D^{BG} is linearly mapped to a range of 0 to 255 in the illustration to avoid the negative value. Therefore, the darker area indicates I_n^{BG} is much larger than I_f^{BG}. As different from the real face and the two mask attacks, the real background is blocked in the image with flash for the photo and video attacks. The values of their D^{BG} are much larger than the ones without flash, *i.e.* their M_BIC values are larger. On the other hand, the real face and the two mask attacks have close M_BIC values because their backgrounds are real and the effect of flash on them is quite similar.

4) Standard Deviation of Background Intensity Change (SD_BIC) Descriptor: As different from the photo and video attacks mentioned in the previous section, the actual background is not covered since only the region of a subject's head is used in the 2D mask attack or curved mask. The light diffusion of masks is different from the one of real face due to the texture and the shape. The light intensity of I_f^{BG} of legitimate and malicious users is different. The variation of the light intensity is measured by

$$\mathrm{SD\_BIC} = \sigma_{D^{BG}} = \sqrt{\frac{\sum_{i=1}^{N}(D^{BG}(x, y) - \mu_{D^{BG}})^2}{N-1}}. \quad (6)$$

Figure 3 shows SD_BIC values of the real face is smaller than that of the mask attacks. It is because the light diffusion of the mask is larger than that of the face. Moreover, the hands captured in the curved mask attack also increase its SD_BIC. Due to a 2D planar structure of an iPad and a photo, flash increases the intensity of the background region uniformly in iPad and paper photo attack, *i.e.* SD_BICs of these attacks are relatively smaller than the ones not covering the real background.

B. Conceptual Discussion

Assume $I(x, y)$ denotes the intensity or grayscale value of the pixel (x, y), where $I(x, y) \in \mathbb{Z}$ is in $[0, 255]$. The intensity of the image without flash (I_n) is defined in (7) according to the Lambertian reflectance law [53]

$$I_n(x, y) = K L_a, \tag{7}$$

where $K \in (0, 1)$ denotes a surface reflectivity at pixel (x, y). Larger K indicates more intensive light is reflected from the surface. $L_a \in (0, \infty)$ is the intensity of the ambient illumination. $L_a = 0$ indicates the dark environment. The model assumes only the ambient light is considered and the intensity of the ambient light is a constant at any point and direction. Therefore, without any additional lighting, as L_a is the same for any object in the same environment, only K is useful for the face liveness detection, *i.e.* the smoothness of a human skin and that of a fake one displayed on 2D planar material are different. However, a face liveness detection only considering K is sensitive to the quality of images and the change of illumination, which has been shown by experiments in the previous study [54].

Based on the Lambertian reflectance law, one additional component is added to the intensity of the image with flash (I_f) defined in (8). In order to make a difference between the scaler and vector multiplication, we omit the dot of the scaler multiplication in these two equations.

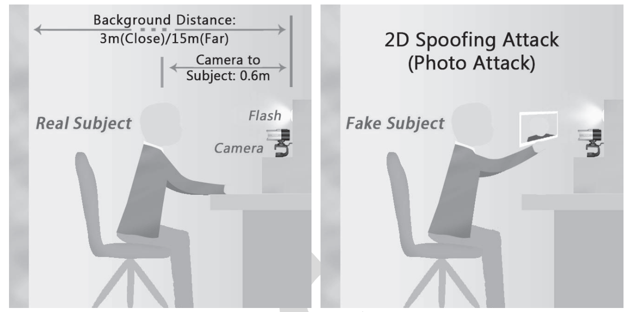
$$I_f(x, y) = K L_a + K L_f \frac{\mathbf{N} \cdot \mathbf{T}}{r^2} = K L_a + K L_f \frac{\cos\theta}{r^2}, \tag{8}$$

where $L_f \in (0, \infty)$ denotes the intensity of the flash. $\mathbf{N}$ is the normal vector to the object surface and $\mathbf{T}$ represents a normalized light-direction vector, pointing from the object surface to the source of flash. θ denotes the angle between $\mathbf{N}$ and $\mathbf{T}$, $\theta \in [0, 90°]$. r is the distance between the flash and the point of the surface. θ as well as r, and $I_f(x, y)$ are inversely proportional, *i.e.* larger θ or r decreases $I_f(x, y)$.

Under the same lighting condition (*i.e.* L_a and L_f are fixed), θ and r of subjects are different due to their shapes. As a result, not only the texture information but also the structure information will be measured. In our proposed model, the LBP_FI descriptor captures the texture information, while SD_FIC, M_BIC and SD_BIC measure the structure information. As a result, the second term of (8) provides extra information to separate the legitimate users from the 2D spoofing attack. It explains why our method may be more accurate than the ones without flash. In addition, more stable liveness detection is expected because of flash, which has a relatively strong illumination in comparison with the ambient light, and it reduces the influence of ambient illumination.

IV. Discussion on Experimental Results

In this section, the performance of our proposed face liveness detection method to encounter different 2D spoofing attacks is evaluated and compared with existing methods experimentally using the dataset we collected under different scenarios. The procedure of the dataset preparation is described at the beginning. Then, the experimental settings as well as the evaluation criterion are introduced. Finally, the experimental results are given and discussed.



(a) (b)

Fig. 4. Settings of sample collection for our dataset: (a) A real subject; (b) A fake subject under photo attack.

(a) (b) (c) (d)

Fig. 5. Examples of the collected images with different distances under normal and uneven ambient illuminance: (a) Far distance (15m) under normal illuminance; (b) Far distance (15m) under uneven illuminance; (c) Close distance (3m) under normal illuminance; (d) Close distance (3m) under uneven illuminance.

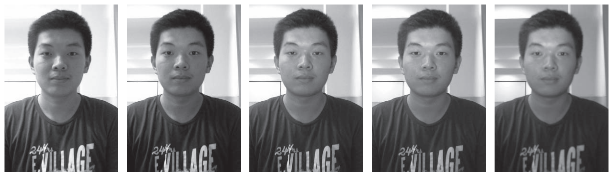
A. Dataset Collection

The dataset[1] for the face liveness detection containing 50 subjects is collected in this paper. The group of subjects consists of 42 male and 8 female with the age from 18 to 21. Each subject is required to sit in front of a web camera (*i.e.* Microsoft Lifecam Studio [55]). Two images, one with flash and another without flash, are taken within a second. Images with 240×360 px are captured, and the face region is around 100×100 px. The detailed setting of the sample collection is illustrated in figure 4.

The distance between a subject and the camera is 0.6m. The flash is set up right above the camera. The distance between the subject and the background is set at 3m and 15m respectively in order to investigate how the distance to background affects the accuracy of liveness detection. The uneven illumination condition, *e.g.* the recognition system is next to a window, is also simulated. A lamp is placed by the side of the subject to create the unbalanced lighting environment. The images with different distances to the background and illumination conditions are shown in figure 5.

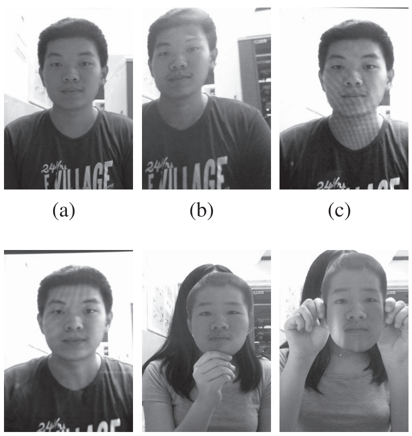
We use illuminance, defined as the total luminous flux incident on a surface per unit area, to represent the intensity of light. Illuminance measures how much incident light illuminates the surface. The only ambient light source in the room in the experiment is ceiling lighting. The illuminance meter is put on the top of the face of a subject, which is

[1] http://www.mlclab.org/dataset/FaceLiveFlash.htm



(a) (b) (c) (d) (e)

Fig. 6. Examples of collected images with additional illuminance values of the target: (a) No extra light; (b) +40lx; (c) +80lx; (d) +120lx; (e) +160lx.



(a) (b) (c)

(d) (e) (f)

Fig. 7. Examples of real face and different types of attacks: (a) Real face; (b) Paper photo attack; (c) iPad photo attack; (d) Video attack; (e) 2D mask attack; (f) Curved mask attack.

parallel to the light source on the ceiling. Without additional device, the natural lighting of the subject is approximately equal to 40lx. To avoid the discomfort to human eyes, we limit the intensity of flash in our proposed method. Four different intensity levels of flash are set to increase the illuminance of the subject by +40lx, +80lx, +120lx, and +160lx. The maximum illuminance adopted by our method, which is 200lx (*i.e.* 40 + 160lx) at 0.6m, is much less than the flash for the camera. For example, the illuminance of the flash for Sony cameras HVL-F60M [56] and HVL-F43RM [57] are approximately 600lx and 400lx respectively at 0.5m. These ensure that the proposed method is practical and the intensity of flash is within the endurance of human eyes. Images with different illuminance values are illustrated in figure 6.

We simulate five different types of 2D spoofing attacks for each person: 1) the photo attack on the A4 sized photographic paper (paper photo attack), 2) the photo attack displayed on iPad with 1024×768 px screen (iPad photo attack), 3) a video (30 fps) being played on iPad with 1024×768 px screen (video attack), 4) the 2D mask attack with the background cut out (2D mask attack), and 5) the curved mask attack with the background cut out (curved mask attack). The examples of a real person and his/her 2D spoofing attacks are shown in figure 7.

For the legitimate user, 2D mask attack and curved mask attack, by considering the distance between the background and the subject, the ambient illumination, and flash illumination, 20 different photos are taken for each person. A total of 1000 samples are collected for each of these classes. Differently, for paper photo attack, iPad photo attack and video attack, the distance between the background and the subject is not considered since the real background cannot be captured. As a result, only 500 images are collected for each of them.

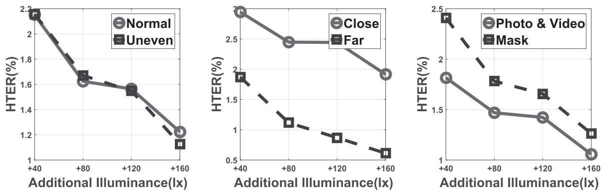
In addition, one thermal image method, which is a hardware based method, is also considered in the experiments. Additional thermal images are collected from 21 subjects by the thermal camera called Seek Thermal Compact XR [58] on a smartphone. The spectral range of the thermal camera is from 7.5 to 14 microns, with 206×156 px image resolution. The low-quality thermal camera is considered since its price is much lower than the professional ones. Therefore, it is more likely to be widely adopted in practice. The factors of environmental illumination and background distance are neglected since they do not affect the decision of a thermal image method. As a result, a total of 126 thermal images were taken, including 21 real face and 105 2D spoofing attack samples.

Temperature of a subject in the real face samples is 33 - 35 °C. As for a paper photo, which is used in 2D mask and curved mask attack, the temperature of a subject in these attacks is 28 - 30 °C, while the one in iPad photo attack is 30 - 32 °C. To evaluate the robustness of the thermal image method, the attack samples are camouflaged by increasing the temperature of 2D spoofing attack. A hot object (*i.e.* a heat pack) is put on the top of the papers, the iPads, and the masks used in the 2D spoofing attack before these objects are put in front of the camera, in order to increase the temperature by 2 - 4 °C. As a result, the temperature difference between a real face and the attack is reduced.

B. Experimental Setting and Evaluation Criterion

The experiments are performed on a computer with 8GB of memory and one Intel processor with i5-4210U cores at 2.40 GHz. A Support Vector Machine (SVM) with the Gaussian kernel implemented by libSVM [59] is applied as the classifier in the experiments. The parameter selection of the penalty coefficient C and Kernel radius γ follow the method of five-fold cross validation using training set based on grid search, which maximizes the classification accuracy. Six methods are selected from different categories of the existing face liveness detection to compare with our proposed method: 1) Traditional LBP method (LBP) [34] in texture-based methods, 2) Eye blinking detection method (EB) [24] in liveness-sign-based methods, 3) Optical Flow Field method (OFF) [22], 4) Diffusion Speed method (DS) [18] in 3D-structure-information-based methods, 5) DMD-LBP-SVM method (DLS) [28] in hybrid methods, and 6) thermal image (TI) in hardware-based methods. A preliminary evaluation is run to tune the parameters of all methods aiming to maximize their average accuracies.

For each experiment, the five-fold cross validation is applied. The performances of the liveness detection methods are evaluated by the running time and also a commonly used criterion, Half Total Error Rate (HTER). HTER is half of False Rejection Rate (FRR) and False Acceptance Rate (FAR), which are both determined by a threshold τ. FRR and FAR are monotonic increasing and decreasing



(a) (b) (c)

Fig. 8. The change of average HTER (%) of the proposed method under different settings and attack types: (a) Under normal and uneven illuminance; (b) Under close and far background distance; (c) Under photo & video and mask attacks.

functions of τ respectively. Larger τ indicates that there is a less probability that a spoof face is misclassified as a live one, and vice versa. When τ is set to the point where FRR and FAR are equal, HTER reaches its minimum. For a dataset $\mathcal{D}$, HTER is defined by

$$HTER(\tau, \mathcal{D}) = \frac{FRR(\tau, \mathcal{D}) + FAR(\tau, \mathcal{D})}{2}, \quad (9)$$

where the range of HTER is from 0 to 1. Lower HTER indicates that the system performs better.

C. Results and Discussion

In this section, we first discuss how the illuminance of the flash affects the performance of our method. Then the proposed model is compared with the existing methods in different scenarios, *i.e.* normal and uneven illumination, the distance between the subject and background, the quality of images and the computational complexity. The discriminate ability of descriptors used in our method is also evaluated. Finally, the performance of the proposed method with the partial knowledge on the type of attacks is discussed.

1) Proposed Method With Different Flash Light Illuminance: This section evaluates how the parameter, the additional illuminance value on the subject increased by flash, affects the performance of the proposed model in different environmental conditions. For each illuminance value and environmental setting, an SVM classifier is trained to distinguish the legitimate users from one type of 2D spoofing attacks. The average performance of the proposed model in different scenarios such as the normal and uneven ambient illuminance, close and far background distance, and photo & video and mask attacks are shown in figure 8. The x-axis and y-axis of the figures represent the additional illuminance values on the subject caused by flash and the average HTER respectively.

In all cases, the values of HTER of the proposed model decreases with the increase of the additional illuminance on the subject. There is no noticeable difference on the increase rates in normal and uneven ambient illumination since flash reduces the influence of the uneven ambient to the detection. However, as the difference between a subject and a background increases by flash, HTER drops more gently in the close distance scenario than the ones in the far distance scenario. As mentioned, detection on mask attacks is more difficult than photo and video attacks since the real background is not blocked by mask attacks. By increasing illuminance, more detail of a mask can be captured. This information is useful to distinguish a mask from a real face. That is why the improvement in the detection of the mask attacks is more significant than that of photo and video attacks.

The results suggest that using a flash light is useful to distinguish 2D spoofing attacks from the legitimate users. Moreover, flash with higher intensity improves the accuracy of the proposed model. This finding is consistent with our explanation of adding flash light in our model in Section III-B. On the other hand, strong flash light will cause the eyes of the users uncomfortable. This parameter is a trade-off between the effectiveness of the liveness detection system and its user friendliness. Two flash settings, *i.e.* +120lx and +160lx shown in figures 6d and 6e, are chosen for the comparison experiments in Sec IV-C.2 and Sect. IV-C.3 to illustrate the performance of our methods using different settings.

2) Comparison With Existing Methods Under Different Attacks: Our proposed methods with +120lx and +160lx, and the five software-based face liveness detection methods, including Traditional LBP method (LBP), Eye blinking detection method (EB), Optical Flow Field method (OFF), Diffusion Speed method (DS), DMD-LBP-SVM method (DLS), and one hardware-based method, *i.e.* thermal image (TI), are evaluated under the 2D spoofing attacks in different environmental settings.

The Student's t-test is conducted to evaluate the confidence level on the difference between the performance of our methods and others. The values of HTER of these experimental results are shown in Table II.

The experimental results indicate that the proposed method with +160lx has the lowest HTER under any type of attack. Moreover, most of the results show that the difference of our method with +160lx and others is statistically significant. On the other hand, our method with +120lx is slightly worse than the one with +160lx in general. These results are consistent with the previous section. Although a soft flash is used, the method with +120lx is still better than the comparison methods in most cases. The results suggest that the use of the flash light improves the 2D spoofing attack detection. The intensity of flash is an important parameter which significantly affects the accuracy of our method.

The proposed method with +160lx is statistically more significant than others in normal illumination with 95% confidence. Although the uneven illumination downgrades the performance of all methods, both of our methods obtain lower HTER in comparison with other methods, except the method with +120lx under iPad photo and 2D mask in the close background distance setting. It indicates that our model is robust in different ambient illuminations. One possible explanation is that the influence of the ambient illumination is reduced since the illuminance of the additional flash light is much stronger. In contrast, the EB method is the most sensitive to the ambient illumination change since the detection of eye blinking requires a clear image of the eyes.

Since EB, OFF and DLS methods only rely on the face region, their performances are independent of the distance between the subject and the background. HTER of all methods with far background distance are generally lower than the

TABLE II

AVERAGE HTER (%) OF THE PROPOSED MODELS WITH +120lx AND +160lx, AND THE COMPARISON METHODS IN DIFFERENT ENVIRONMENTAL SETTINGS (N: NORMAL AMBIENT ILLUMINANCE, U: UNEVEN AMBIENT ILLUMINANCE, C: CLOSE BACKGROUND DISTANCE, F: FAR BACKGROUND DISTANCE, AVG: AVERAGE HTER OF ALL SETTINGS)

Methods	Attack Types and Settings														AVG
	Photo(Paper)		Photo(iPad)		Video		2D Mask				Curved Mask				
	N	U	N	U	N	U	C+N	C+U	F+N	F+U	C+N	C+U	F+N	F+U	
Proposed (+120lx)	1.23	1.36	0.83	1.42	1.95	1.74	3.79	1.90	0.99	**0.68**	1.84	2.26	**0.33**	1.47	1.56
Proposed (+160lx)	**1.03**	**1.13**	**0.50**	**0.66**	1.07	1.95	3.90	**1.12**	**0.35**	0.88	**1.01**	1.62	0.70	**0.51**	**1.17**
EB$^{\wedge}$	10.21*◊	7.95*◊	9.97*◊	8.09*◊	17.56*◊	9.93*◊	12.16*◊	8.33*◊	12.16*◊	8.33*◊	15.57*◊	10.36*◊	15.57*◊	10.36*◊	11.18
OFF$^{\wedge}$	9.97*◊	6.17*◊	4.94*◊	5.15*◊	12.23*◊	11.00*◊	2.06*◊	2.74◊	2.06*◊	2.74	6.16*◊	2.83*◊	2.15*◊	2.83*◊	5.24
DLS$^{\wedge}$	2.88*◊	4.52*◊	1.85*◊	2.72*◊	3.64*◊	5.42*◊	2.78*	4.54*◊	2.78*◊	4.54*◊	5.78*◊	5.40*◊	5.78*◊	5.40*◊	4.15
TI$^{\circ}$	1.19	1.19◊	2.47*◊	2.47*◊	3.66*◊	3.66*◊	1.19*◊	1.19*◊	1.19*◊	1.19*◊	1.22*◊	**1.22***◊	1.22*◊	1.22*◊	1.73
$TI_{att}^{\circ\triangle}$	1.28*◊	1.28*◊	3.66*◊	3.66*◊	6.04*◊	6.04*◊	5.23*◊	5.23*◊	5.23*◊	5.23*◊	2.25*◊	2.25*◊	2.25*◊	2.25*◊	3.71
LBP	6.33*◊	5.30*◊	2.00	3.25*	3.10*◊	3.83◊	1.40*◊	1.80*◊	1.28◊	1.36*	1.68*	3.12*◊	0.66	3.08	2.73
DS	1.80*◊	2.70*◊	1.67*◊	1.17*◊	4.10*◊	5.73*◊	5.20*◊	2.14*◊	4.13*◊	3.69*◊	1.36*◊	2.35*◊	1.20*◊	2.40*◊	2.83
LBP_F (+120lx)	3.27*◊	1.25*	1.25*◊	6.67*◊	2.53	3.87*◊	3.82*◊	5.75*◊	1.00*◊	2.50*◊	**1.01***	3.04	5.25*◊	1.75*◊	3.07
LBP_F (+160lx)	2.02*◊	1.76	6.51*◊	2.26*◊	1.75*◊	2.28*◊	2.53*◊	5.12*◊	1.25	6.05*◊	4.60	5.62	4.86*◊	2.00*◊	3.47
DS_F (+120lx)	4.04*◊	4.59*◊	4.28*◊	1.27	2.51◊	1.51	4.56*◊	3.53*◊	7.34*◊	2.51*◊	1.76*◊	5.29*◊	4.58*◊	2.78*◊	3.61
DS_F (+160lx)	6.09*◊	5.08*◊	3.04*◊	2.51*◊	1.52*	1.77	6.30*◊	3.55*◊	5.56*◊	2.79*◊	6.06*◊	4.29*◊	7.10*◊	5.29*◊	4.35
LBP+LBP_F (+120lx)	5.29*◊	3.97*◊	1.67*	2.48*◊	2.60*◊	3.64*◊	2.96◊	1.64*◊	1.25*◊	1.75*◊	1.86◊	3.02*◊	0.69◊	1.72◊	2.47
LBP+LBP_F (+160lx)	4.99*◊	3.58*◊	1.86*◊	3.27*◊	2.47*◊	3.44*◊	2.10*◊	1.97*◊	0.99*◊	1.15*	4.47*◊	3.18*◊	0.55◊	1.34◊	2.53
DS+DS_F (+120lx)	3.03*◊	2.79*◊	2.77*◊	1.51*◊	**0.75**	**1.01***	2.51◊	2.78*	2.78*◊	2.51*◊	2.01*	2.26◊	5.05*◊	2.00*◊	2.41
DS+DS_F (+160lx)	2.54*◊	1.53*◊	1.26*◊	1.51*◊	1.76	1.25*	**1.00***◊	3.02*◊	1.26*◊	2.26*◊	1.51*	2.02*◊	2.53*◊	3.80*◊	1.95

◊ Statistically significant difference with 95% confidence in comparison with our proposed method (+120lx) using the Student's t-test.
* Statistically significant difference with 95% confidence in comparison with our proposed method (+160lx) using the Student's t-test.
$^{\wedge}$ The method is independent to background distance.
$^{\circ}$ The method is independent to both background distance and environmental illuminance.
$^{\triangle}$ The temperature of the 2D spoofing attack is raised intensionally in this method.

ones with close background distance. It is because the depth information is more easily detected with the increase of the background distance. In both scenarios, the proposed models maintain stable and satisfying performance.

The significant temperature difference between a real face and the spoofing attacks causes TI to achieve a satisfying performance and the result is more accurate than other existing face liveness detection methods. However, HTER of TI is still lower than the one for our proposed methods. Moreover, if an adversary raises the temperature of the object in order to reduce the difference between a real face and the attack, HTER of TI increases dramatically. The results are shown in the row of TI_{att} in Table II. It indicates a security hole of TI which should be further studied to increase its robustness in an adversarial environment.

We further investigate whether or not the use of flash image will improve the accuracy of a face liveness detection method. HTER of LBP and DS are compared with the one of LBP and DS on flash images (LBP_F (+120lx), LBP_F (+160lx), DS_F (+120lx), and DS_F (+160lx)), combination of LBP and LBP_F with average fusion (LBP+LBP_F (+120lx), LBP+LBP_F (+160lx)) with average fusion, and combination of DS and DS_F with average fusion (DS+DS_F (+120lx), DS+DS_F (+160lx)) in Table II.

The experimental results show that the method using only flash images is not consistently better the one with non-flash images. For LBP, flash images improve the detection of photo and video attacks, *i.e.* the average HTER on photo and video attacks of LBP_F is lower than 1.46 under normal ambient illuminance. However, LBP with flash images becomes less accurate on 2D and curved mask attacks than LBP with non-flash images. In 8 out of 14 cases, LBP_F with +120lx and +160lx flash images is better than LBP. It is 7 out of 14 cases for LBP_F with +160lx flash images. However, the average HTER of LBP (2.73) is slightly lower than the one of LBP_F (3.07 for +120 and 3.47 for +160). This indicates that LBP with flash images is not robust consistently, which explains why additional structure features are considered in our proposed method. For DS, the contribution of flash images is less insignificant. Only 3 out of 14 cases and 1 out of 14 cases show that DS_F (+120lx) and DS_F (+160lx) are better than DS with 95% significant confidence. This may be because DS focuses on weak light diffusion on a human face, which becomes difficult to capture with flash.

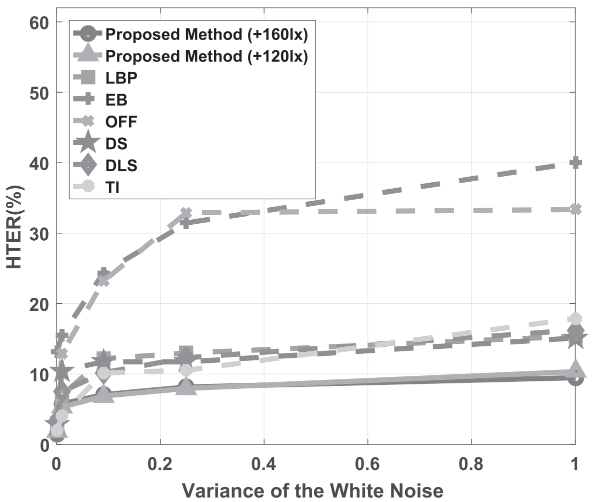


Fig. 9. Average HTER (%) of the proposed methods with +120lx and +160lx, and the comparison methods on images contaminated by the white noise with difference variances.

While considering the fusion of the methods with flash and non-flash images, HTER of LBP+LBP_F and DS+DS_F is significantly lower than the one for LBP, DS, LBP_F and DS_F generally. The results suggest the importance of considering both flash and non-flash images. The utilization of both images may provide a useful comparison to indicate whether the subject is from a spoofing attack. Although LBP+LBP_F and DS+DS_F achieves relatively good performance, their HTER is higher than the one for our methods (both with +120lx and +160lx) in all cases except video attack under normal and uneven illuminance, and 2D mask under illuminance for close background distance.

In summary, the experimental results demonstrate that the proposed method with +160lx successfully outperforms other comparison methods under various types of spoofing attacks. Although a flash with lower intensity is used, our method with +120lx still achieves satisfying results which are better than other methods generally. The performance of our method is also less sensitive to different environmental factors including the background distance and the ambient illuminance.

3) Comparison With Existing Methods With Noisy Images: The robustness of the face liveness detection to noisy images is evaluated. Only the close background distance and normal ambient illuminance are considered in this comparison. The average HTER of the detection method for all five types of attacks is calculated. All detection methods are trained with untainted training set. The white Gaussian noise with the variance = 0.01, 0.09, 0.25 and 1, and the mean = 0 are added to each testing sample, which has been normalized to the interval [0, 1]. The examples of the noisy images are shown in figure 10.

The experimental results shown in figure 9 suggest that the performances of all methods suffer from the noise, *i.e.* HTER increases with the noise. There is no significant difference between the performance of our methods with +120lx and +160lx on images with the white noise with different variances. They achieve the most robust performance among all methods. As real faces and 2D spoofing attacks are clearly

(a) (b) (c) (d) (e)

Fig. 10. Examples of noise images with different variances: (a) 0; (b) 0.01; (c) 0.09; (d) 0.25; (e) 1.

TABLE III
AVERAGE RUNNING TIME OF FEATURE EXTRACTION AND CLASSIFICATION FOR THE PROPOSED METHOD AND COMPARISON METHODS

	Our Method	LBP	EB	OFF	DS	DLS	TI
Feature Extraction(s)	**0.04**	0.17	22.93	1.71	0.16	43.13	0.07
Classification(s)	0.20	0.19	**0.01**	**0.01**	0.16	0.19	0.20
Total Time(s)	**0.24**	0.36	22.94	1.78	0.32	43.32	0.27

separated in our system, the white noise with large variance does not affect our results dramatically, *i.e.* their HTER values increase slightly with the increase of variance of the white noise. The results indicate that our model is robust to the while noise even though only flash with weak intensity is used. HTER of TI, DS, DLS and LBP increases more slowly than EB and OFF. Since EB and OFF highly depend on pixel-level analysis, they are noise sensitive. This observation agrees with the result in the previous section.

4) Computational Complexity: The computational complexity of the methods in terms of the average running time of feature extraction and the classification are given in Table III. The proposed method has the lowest computational complexity of feature extraction since only LBP_FI as well as the standard deviation and mean values are extracted. As different from the traditional LBP method which extracts the value from the whole picture, LBP_FI of our model only measures the face region which is much smaller than the original image. EB and DLS cost the most extraction time because complicated features are extracted from hundreds of frames. As the extraction of LBP and intensity histogram is required for TI, its time complexity is slightly larger than the one of the proposed method. The classification times of all methods are similar except EB and OFF since they only consider a single, one-dimensional feature. In conclusion, although two images are processed in our method, its time complexity is still relatively low in comparison with other detection methods.

5) Effectiveness of the Descriptors of the Proposed Method: The discriminant ability of the descriptors in the proposed method is evaluated in this section. A classifier is trained using one combination of features each time. The settings such as the close background distance, normal ambient illuminance and +120lx additional illuminance are considered in this experiment. From the results given in Table IV, LBP_FI is the most critical descriptor which affects the performance of

TABLE IV

AVERAGE HTER (%) OF THE PROPOSED MODEL WITH +120lx AND DIFFERENT FEATURE COMBINATIONS IN THE CLOSE BACKGROUND DISTANCE AND NORMAL AMBIENT ILLUMINANCE. DESCRIPTOR: ① LBP_FI, ② SD_FIC, ③ M_BIC AND ④ SD_BIC

Attack Type	Feature Combinations														
	①	②	③	④	①②	①③	①④	②③	②④	③④	①②③	①②④	①③④	②③④	①②③④
Photo (Paper)	2.50	20.29	9.48	23.86	1.91	2.50	1.67	11.17	16.71	10.17	1.67	2.50	2.67	10.83	**1.62**
Photo (iPad)	1.83	15.83	3.33	17.47	2.67	2.67	2.92	1.83	5.41	1.94	1.67	2.5	1.74	2.50	**0.84**
Video	1.83	11.17	2.67	12.30	1.91	1.67	1.83	1.67	6.17	2.50	1.83	1.67	1.83	1.00	**0.83**
2D Mask	4.24	17.91	26.62	24.66	6.00	4.33	6.35	15.73	7.89	18.67	6.66	5.17	5.72	6.89	**3.10**
Curved Mask	4.33	17.93	25.58	24.67	5.50	3.91	4.50	16.39	14.39	13.11	2.91	2.65	2.58	11.02	**1.91**

TABLE V

AVERAGE HTER (%) OF THE PROPOSED MODELS WITH +120lx USING LBP_FI, DS_FI AND DoG_FI IN THE CLOSE BACKGROUND DISTANCE AND NORMAL AMBIENT ILLUMINANCE DESCRIPTOR: ① LBP_FI, ② SD_FIC, ③ M_BIC AND ④ SD_BIC, ⑤ DS_FI, ⑥ DoG_FI

Attack Type	Feature Combinations		
	①+②③④	⑤+②③④	⑥+②③④
Photo (Paper)	1.23	**1.08***	2.69*
Photo (iPad)	**0.83**	0.85*	1.17
Video	1.95	2.18	**1.25***
2D Mask	3.79	**3.23***	3.28
Curved Mask	1.84	1.81*	**1.77**
Average	1.93	**1.83**	2.03

* Statistically significant difference with 95% confidence in comparison with our proposed method (+120lx) using the Student's t-test.

TABLE VI

AVERAGE HTER (%) OF OUR METHOD WITH +120lx TRAINED WITH DIFFERENT KINDS OF ATTACKS

Training \ Test	Paper Photo	iPad Photo	Video	2D Mask	Curved Mask	All
Photo (Paper)	**1.23**	2.73	2.73	7.31	8.27	10.71
Photo (iPad)	1.77	**0.83**	0.97	4.59	9.13	9.28
Video	2.59	2.63	**1.95**	5.45	8.01	8.42
2D Mask	2.68	4.45	4.45	**3.79**	5.56	5.19
Curved Mask	1.86	2.68	2.59	1.86	**1.84**	4.51
All	0.95	**0.00**	0.92	0.86	0.85	**0.00**

the proposed model significantly. HTER of classifiers with any combination containing LBP_FI is lower than 6.7%. Furthermore, M_BIC also plays an important role in detecting attack of iPad and video where HTER of the classifiers with any combination of M_BIC is lower than 3.33%. It may be because the severe reflection of an iPad screen increases the mean value of the background region, which makes these two types of attacks more differentiable from normal faces. The descriptors SD_FIC and SD_BIC perform badly individually. For instance, the HTER of using only SD_FIC and SD_BIC is larger than 11% and 12% respectively for all attacks. However, HTER of our model using all descriptors is the lowest in each row, which suggests that although an individual descriptor may not perform well, it works well with other descriptors as a group and every one of them has a positive impact on the 2D spoofing attack detection.

Our model is also evaluated using other descriptors. LBP, which plays a key role in our model, is replaced by more advanced features, *i.e.* DS [18] and DoG [14]. Similar to LBP_FI described in Sec III-A.1, DS and DoG are applied to the image with flash in our model, named DS_FI and DoG_FI. Table V shows HTER of our original model, and our revised models in which LBP_FI is replaced by DS_FI and DoG_FI.

As DS focuses on the structure difference of the subject's face, our method using DS_FI has more satisfying performance under paper photo and 2D mask attacks than our original method. However, our original model achieves lower HTER than DS_FI in other attacks. On the other hand, the models using LBP_FI are better in photo attacks but worse in video and mask attacks than the ones using DoG_FI. The difference on HTER of the models using LBP_FI, DS_FI, and DoG_FI is less than 1%, *i.e.* they have similar performance. However, by considering its short feature extraction time, LBP_FI is a suitable feature for our model.

6) Partial Knowledge on the Attack Types: The face liveness detection may be invaded by an unseen attack in reality. In this section, we assume that the defenders know a 2D spoofing attack is used but not the type. The proposed method is trained by one of the attacks and then is evaluated by another. We consider the scenario with the close background distance and normal environmental illuminance. +120lx additional illuminance is used in our model. The results are displayed in Table VI. Each row represents our method trained by one type of attack while each column is the evaluation using the test set with another type of attack. When all types of attacks are used in the training (test) phase, the row and the column are named by "All".

The performance of our method drops when the training and test set contain different types of attacks. The five 2D spoofing attacks applied in the experiment can be categorized into two types: 1) photo & video attack, and 2) mask attack. When the attacks in the training and test set are in the same category, our method maintains a good performance. However, HTER of our model is larger when the training and test set are different, except 2D mask attack. For example, for the model using a training set with paper photo attack, its HTER on the test set with iPad photo attack (2.73%) is much lower than the one with 2D mask attack (7.31%). The classifier using a training set with 2D mask attack detects paper photo attack more accurately than 2D mask attack in the test phase. This is mainly because paper photo attack is similar to 2D mask attack but easier to be identified. This observation in general agrees with other classification problems, namely, the similarity between training and test sets affects the performance of detection.

When the proposed method is trained by using all kinds of attacks, the performance of classifying each attack is satisfying, which is slightly worse than the one trained with the same attack. Moreover, the HTER value of classifying all attacks is 0.0%, which is the lowest value among all methods trained with one attack. This result demonstrates that our method can handle a complicated situation arising from several kinds of attacks. If all kinds of 2D spoofing attacks are obtained in advance, our method can protect the system effectively.

V. Conclusion and Future Work

A face liveness detection method against 2D spoofing attack using flash is proposed in this paper. The descriptors of the texture (*i.e.* LBP_FI) and structure analysis (*i.e.* SD_FIC, M_BIC and SD_BIC) are carefully designed to capture the difference from two images of the subject, one with flash and the other without flash. Our method has satisfying performance because flash enhances the differences between legitimate users and attacks. The conceptual discussion is also given based on the Lambertian reflectance law. In contrast to the existing methods, the proposed model combines the advantage of the software and hardware approaches which are high accuracy, high robustness, low computational complexity and low setup cost.

A dataset containing 50 subjects with 2D spoofing attacks, including paper photo, iPad photo, video, 2D mask and curved mask attack, are collected. In order to compare with the thermal image method, thermal images of 21 subjects with real and five types of attacks are also collected. Our method is also compared experimentally with five software-based and one hardware-based liveness detection methods. The experimental results show that the proposed method is better in terms of accuracy and running time. In addition, the robustness of our method to noisy images and different environmental settings including the background distance and ambient illuminance is better than other methods.

The tradeoff of the superiority of our method is the installation of an additional hardware, *i.e.* flash. It may limit the applications of our method, *e.g.* frontal flash is not a necessary device for a smartphone. However, different from other hardware-based methods, it may not be a serious issue since the installation cost of a flash is low in comparison with other hardware used, *e.g.* a thermal camera. Moreover, flash becomes more popular and can be found in many systems recently, *e.g.* frontal flash is more popular recently due to the popularity of the selfie.

Although the illuminance of flash in our current model is no harm to human eyes and it is also much lower than the illuminance of flash used in a camera, user comfort is a concern. A possible solution to overcome this limitation is to adjust the angle of flash on a subject. If flash is not installed at the eye level, the lighting of flash will not directly irritate human eyes and a subject will feel more comfortable. The angle of flash should be determined according to not only the detection accuracy but also installation difficulty. Other robust features may be considered in our model due to the change of lighting angle.

With the promising results obtained in this study of using flash in against 2D spoofing attack, one possible future work is to focus on exploring the performance of the proposed model on the detection of more advanced attacks, such as, the 3D spoofing attacks, for instance, rigid 3D mask and 3D face models with various expressions. The reflected light from a real face and a 3D mask is expected to be different since they have different surface reflectivity. Moreover, the texture detail of the 3D masks may also be enhanced by the flash. As a result, the additional lighting should be useful to separate legitimate users from the attacks if suitable descriptors can be identified.

References

[1] A. K. Jain, A. Ross, and S. Prabhakar, "An introduction to biometric recognition," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 14, no. 1, pp. 4–20, Jan. 2004.

[2] U. Uludag, S. Pankanti, S. Prabhakar, and A. K. Jain, "Biometric cryptosystems: Issues and challenges," *Proc. IEEE*, vol. 92, no. 6, pp. 948–960, Jun. 2004.

[3] W. Zhao, R. Chellappa, P. J. Phillips, and A. Rosenfeld, "Face recognition: A literature survey," *ACM Comput. Surv.*, vol. 35, no. 4, pp. 399–458, 2003.

[4] A. F. Abate, M. Nappi, D. Riccio, and G. Sabatino, "2D and 3D face recognition: A survey," *Pattern Recognit. Lett.*, vol. 28, no. 14, pp. 1885–1906, Oct. 2007.

[5] A. Wagner, J. Wright, A. Ganesh, Z. Zhou, H. Mobahi, and Y. Ma, "Toward a practical face recognition system: Robust alignment and illumination by sparse representation," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 34, no. 2, pp. 372–386, Feb. 2012.

[6] C.-P. Wei, C.-F. Chen, and Y.-C. F. Wang, "Robust face recognition with structurally incoherent low-rank matrix decomposition," *IEEE Trans. Image Process.*, vol. 23, no. 8, pp. 3294–3307, Aug. 2014.

[7] S. Gao, Y. Zhang, K. Jia, J. Lu, and Y. Zhang, "Single sample face recognition via learning deep supervised autoencoders," *IEEE Trans. Inf. Forensics Security*, vol. 10, no. 10, pp. 2108–2118, Oct. 2015.

[8] I. Chingovska, A. R. dos Anjos, and S. Marcel, "Biometrics evaluation under spoofing attacks," *IEEE Trans. Inf. Forensics Security*, vol. 9, no. 12, pp. 2264–2276, Dec. 2014.

[9] B. Biggio, Z. Akhtar, G. Fumera, G. L. Marcialis, and F. Roli, "Security evaluation of biometric authentication systems under real spoofing attacks," *IET Biometrics*, vol. 1, no. 1, pp. 11–24, Mar. 2012.

[10] Z. Akhtar, C. Micheloni, and G. L. Foresti, "Biometric liveness detection: Challenges and research opportunities," *IEEE Security Privacy*, vol. 13, no. 5, pp. 63–72, Sep./Oct. 2015.

[11] R. Tronci *et al.*, "Fusion of multiple clues for photo-attack detection in face recognition systems," in *Proc. IEEE Int. Joint Conf. Biometrics (IJCB)*, Washington, DC, USA, Oct. 2011, pp. 1–6.

[12] I. Chingovska *et al.*, "The 2nd competition on counter measures to 2D face spoofing attacks," in *Proc. IAPR Int. Conf. Biometrics (ICB)*, Madrid, Spain, Jun. 2013, pp. 1–6.

[13] A. Anjos and S. Marcel, "Counter-measures to photo attacks in face recognition: A public database and a baseline," in *Proc. IEEE Int. Joint Conf. Biometrics (IJCB)*, Washington, DC, USA, Oct. 2011, pp. 1–7.

[14] X. Tan, Y. Li, J. Liu, and L. Jiang, "Face liveness detection from a single image with sparse low rank bilinear discriminative model," in *Proc. 11th Eur. Conf. Comput. Vis. (ECCV)*, Heraklion, Greece, Sep. 2010, pp. 504–517.

[15] D. Wen, H. Han, and A. K. Jain, "Face spoof detection with image distortion analysis," *IEEE Trans. Inf. Forensics Security*, vol. 10, no. 4, pp. 746–761, Apr. 2015.

[16] Z. Zhang, J. Yan, S. Liu, Z. Lei, D. Yi, and S. Z. Li, "A face antispoofing database with diverse attacks," in *Proc. 5th IAPR Int. Conf. Biometrics (ICB)*, New Delhi, India, Mar./Apr. 2012, pp. 26–31.

[17] I. Chingovska, A. Anjos, and S. Marcel, "On the effectiveness of local binary patterns in face anti-spoofing," in *Proc. Int. Conf. Biometrics Special Interest Group (BIOSIG)*, Darmstadt, Germany, Sep. 2012, pp. 1–7.

[18] W. Kim, S. Suh, and J.-J. Han, "Face liveness detection from a single image via diffusion speed model," *IEEE Trans. Image Process.*, vol. 24, no. 8, pp. 2456–2465, Aug. 2015.

[19] S. Chakraborty and D. Das. (2014). "An overview of face liveness detection." [Online]. Available: https://arxiv.org/abs/1405.2227

[20] J. Li, Y. Wang, T. Tan, and A. K. Jain, "Live face detection based on the analysis of Fourier spectra," *Proc. SPIE*, vol. 5404, pp. 296–303, Apr. 2004. AQ:3

[21] J. Määttä, A. Hadid, and M. Pietikäinen, "Face spoofing detection from single images using micro-texture analysis," in *Proc. Int. Joint Conf. Biometrics (IJCB)*, Washington, DC, USA, Oct. 2011, pp. 1–7.

[22] W. Bao, H. Li, N. Li, and W. Jiang, "A liveness detection method for face recognition based on optical flow field," in *Proc. Int. Conf. Image Anal. Signal Process. (IASP)*, Taizhou, China, Apr. 2009, pp. 233–236.

[23] T. Choudhury, B. Clarkson, T. Jebara, and A. Pentland, "Multimodal person recognition using unconstrained audio and video," in *Proc. Int. Conf. Audio Video-Based Person Authentication*, 1999, pp. 176–181.

[24] J.-W. Li, "Eye blink detection based on multiple Gabor response waves," in *Proc. Int. Conf. Mach. Learn. Cybern. (ICMLC)*, vol. 5. Jul. 2008, pp. 2852–2856.

[25] H. Yu, T.-T. Ng, and Q. Sun, "Recaptured photo detection using specularity distribution," in *Proc. 15th IEEE Int. Conf. Image Process. (ICIP)*, Oct. 2008, pp. 3140–3143.

[26] J. Galbally, S. Marcel, and J. Fierrez, "Image quality assessment for fake biometric detection: Application to iris, fingerprint, and face recognition," *IEEE Trans. Image Process.*, vol. 23, no. 2, pp. 710–724, Feb. 2014.

[27] G. Kim, S. Eum, J. K. Suhr, D. I. Kim, K. R. Park, and J. Kim, "Face liveness detection based on texture and frequency analyses," in *Proc. 5th IAPR Int. Conf. Biometrics (ICB)*, New Delhi, India, Mar./Apr. 2012, pp. 67–72.

[28] S. Tirunagari, N. Poh, D. Windridge, A. Iorliam, N. Suki, and A. Ho, "Detection of face spoofing using visual dynamics," *IEEE Trans. Inf. Forensics Security*, vol. 10, no. 4, pp. 762–777, Apr. 2015.

[29] G. Chetty, "Biometric liveness detection based on cross modal fusion," in *Proc. 12th Int. Conf. Inf. Fusion (FUSION)*. Seattle, WA, USA, Jul. 2009, pp. 2255–2262.

[30] K. Kollreider, H. Fronthaler, and J. Bigun, "Evaluating liveness by face images and the structure tensor," in *Proc. 4th IEEE Workshop Automat. Identificat. Adv. Technol.*, Buffalo, NY, USA, Oct. 2005, pp. 75–80.

[31] A. Pinto, W. R. Schwartz, H. Pedrini, and A. Rocha, "Using visual rhythms for detecting video-based facial spoof attacks," *IEEE Trans. Inf. Forensics Security*, vol. 10, no. 5, pp. 1025–1038, May 2015.

[32] Z. Zhang, D. Yi, Z. Lei, and S. Z. Li, "Face liveness detection by learning multispectral reflectance distributions," in *Proc. FG*, Santa Barbara, CA, USA, Mar. 2011, pp. 436–441.

[33] W. Liu, "Face liveness detection using analysis of Fourier spectra based on hair," in *Proc. Int. Conf. Wavelet Anal. Pattern Recognit. (ICWAPR)*, Lanzhou, China, Jul. 2014, pp. 75–80.

[34] T. Ojala, M. Pietikäinen, and T. Mäenpää, "Multiresolution gray-scale and rotation invariant texture classification with local binary patterns," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 24, no. 7, pp. 971–987, Jul. 2002.

[35] Z. Boulkenafet, J. Komulainen, and A. Hadid, "Face spoofing detection using colour texture analysis," *IEEE Trans. Inf. Forensics Security*, vol. 11, no. 8, pp. 1818–1830, Aug. 2016.

[36] M. Smiatacz, "Liveness measurements using optical flow for biometric person authentication," *Metrol. Meas. Syst.*, vol. 19, no. 2, pp. 257–268, 2012.

[37] A. Anjos, M. M. Chakka, and S. Marcel, "Motion-based countermeasures to photo attacks in face recognition," *IET Biometrics*, vol. 3, no. 3, pp. 147–158, Sep. 2014.

[38] H. K. Jee, S. U. Jung, and J. H. Yoo, "Liveness detection for embedded face recognition system," in *Proc. World Acad. Sci., Eng. Technol.*, Vienna, Austria, Dec. 2006, pp. 29–32.

[39] G. Pan, L. Sun, Z. Wu, and S. Lao, "Eyeblink-based anti-spoofing in face recognition from a generic webcamera," in *Proc. IEEE 11th Int. Conf. Comput. Vis. (ICCV)*, Rio de Janeiro, Brazil, Oct. 2007, pp. 1–8.

[40] G. Chetty and M. Wagner, "Multi-level liveness verification for face-voice biometric authentication," in *Proc. Biometrics Symp., Special Session Res. Biometric Consortium Conf.*, Baltimore, MD, USA, Sep. 2006, pp. 1–6.

[41] K. Patel, H. Han, and A. K. Jain, "Secure face unlock: Spoof detection on smartphones," *IEEE Trans. Inf. Forensics Security*, vol. 11, no. 10, pp. 2268–2283, Oct. 2016.

[42] T. I. Dhamecha, A. Nigam, R. Singh, and M. Vatsa, "Disguise detection and face recognition in visible and thermal spectrums," in *Proc. Int. Conf. Biometrics*, Jun. 2013, pp. 1–8.

[43] A. Lagorio, M. Tistarelli, M. Cadoni, C. Fookes, and S. Sridharan, "Liveness detection based on 3D face shape analysis," in *Proc. 1st Int. Workshop Biometrics Forensics (IWBF)*, Lisbon, Portugal, Apr. 2013, pp. 1–4.

[44] S. Kim, Y. Ban, and S. Lee, "Face liveness detection using a light field camera," *Sensors*, vol. 14, no. 12, pp. 22471–22499, Jan. 2014.

[45] G. Pan, L. Sun, Z. Wu, and Y. Wang, "Monocular camera-based face liveness detection by combining eyeblink and scene context," *Telecommun. Syst.*, vol. 47, pp. 215–225, Aug. 2011.

[46] Y. Kim, J.-H. Yoo, and K. Choi, "A motion and similarity-based fake detection method for biometric face recognition systems," *IEEE Trans. Consum. Electron.*, vol. 57, no. 2, pp. 756–762, May 2011.

[47] D. A. Socolinsky, A. Selinger, and J. D. Neuheisel, "Face recognition with visible and thermal infrared imagery," *Comput. Vis. Image Understand.*, vol. 91, nos. 1–2, pp. 72–114, Jun./Aug. 2003.

[48] G. Chetty and M. Wagner, "Biometric person authentication with liveness detection based on audio-visual fusion," *Int. J. Biometrics*, vol. 1, no. 4, pp. 463–478, 2009.

[49] G. Chetty, "Biometric liveness checking using multimodal fuzzy fusion," in *Proc. IEEE Int. Conf. Fuzzy Syst.*, Barcelona, Spain, Jul. 2010, pp. 1–8.

[50] T. Wang, J. Yang, Z. Lei, S. Liao, and S. Z. Li, "Face liveness detection using 3D structure recovered from a single camera," in *Proc. ICB*, Madrid, Spain, Jun. 2013, pp. 1–6.

[51] M. Nilsson, J. Nordberg, and I. Claesson, "Face detection using local smqt features and split up snow classifier," in *Proc. IEEE Int. Conf. Acoust., Speech, Signal Process. (ICASSP)*, vol. 2. Apr. 2007, pp. II-589–II-592.

[52] T. Ahonen, A. Hadid, and M. Pietikäinen, "Face recognition with local binary patterns," in *Proc. 8th Eur. Conf. Comput. Vis.*, Prague, Czech Republic, May 2004, pp. 469–481.

[53] R. Basri and D. W. Jacobs, "Lambertian reflectance and linear subspaces," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 25, no. 2, pp. 218–233, Feb. 2003.

[54] P. Liu, F. Zafar, and A. Badano, "The effect of ambient illumination on handheld display image quality," *J. Digit. Imag.*, vol. 27, no. 1, pp. 12–18, Feb. 2014.

[55] *Microsoft LifeCam Studio*. [Online]. Available: http://www.microsoftstore.com/store/msusa/en_US/pdp/productID.258411900 AQ:4

[56] *F60M External Flash for Multi-Interface Shoe—HVL-F60M—Sony us*. [Online]. Available: https://www.sony.com/electronics/interchangeable-lens-cameras-flashes-lights/hvl-f60m

[57] *F43M External Flash for Multi-Interface Shoe—HVL-F43M—Sony us*. [Online]. Available: https://www.sony.com/electronics/ interchangeable-lens-cameras-flashes-lights/hvl-f43m

[58] *CompactXR—Seek Thermal*. [Online]. Available: http://www.thermal.com/products/compactxr/

[59] C.-C. Chang and C.-J. Lin, "LIBSVM: A library for support vector machines," *ACM Trans. Intell. Syst. Technol.*, vol. 2, no. 3, pp. 27:1–27:27, 2011.

Patrick P. K. Chan received the Ph.D. degree from The Hong Kong Polytechnic University in 2009. He is currently an Associate Professor with the School of Computer Science and Engineering, and the person in charge of machine learning and the Cybernetics Research Laboratory, South China University of Technology, Guangzhou, China. He is also a part-time Lecturer with the Hyogo College of Medicine, Japan. His current research interests include pattern recognition, multiple classifier system, biometric, computer security, deep learning, and reinforcement learning. He was a member of the governing boards of the IEEE SMC Society from 2014 to 2016. He serves as an Organizing Committee Chair of several international conferences. He was also the Chairman of the IEEE SMCS Hong Kong Chapter 14–15. He is the Counselor of the IEEE Student Branch, South China University of Technology. He is an associate editor for international journals, including *Information Sciences* and the *International Journal of Machine Learning and Cybernetics*.

Weiwen Liu received the B.S. degree in computer science and technology from the South China University of Technology in 2013. She is currently pursuing the Ph.D. degree in computer science and engineering with The Chinese University of Hong Kong. Her research interests include adversarial learning, machine learning, and machine learning algorithms.

Danni Chen received the B.S. degree from the School of Computer Science and Engineering, South China University of Technology, China, in 2016, where she is currently pursuing the M.S. degree. Her current research interests include computer vision and machine learning.

Daniel S. Yeung (F'04) received the Ph.D. degree in applied mathematics from Case Western Reserve University. He was an Assistant Professor of mathematics and computer science with the Rochester Institute of Technology, USA, as a Research Scientist with the General Electric Corporate Research Center, USA, and as a System Integration Engineer with TRW, USA. He was a Visiting Professor with the School of Computer Science and Engineering, South China University of Technology, Guangzhou, China, from 2008 to 2015. His current research interests include neural-network sensitivity analysis, data mining, and big data analytic. He was the Chairman of the Department of Computing, The Hong Kong Polytechnic University, Hong Kong, and a Chair Professor from 1999 to 2006. He is a Past President of the IEEE Systems and the Man and Cybernetics Society. He is a Co-Editor-in-Chief of the Springer *International Journal on Machine Learning and Cybernetics*.

Fei Zhang received the Ph.D. degree from the South China University of Technology, Guangzhou, China. She is currently a Lecturer with the College of Computer and Information Engineering, Henan Normal University, Xinxiang, China. Her current research interests include machine learning, computer security, and recommender system.

Xizhao Wang (M'03–SM'04–F'12) received the Ph.D. degree in computer science from the Harbin Institute of Technology in 1998. From 1998 to 2001, he was with the Department of Computing, The Hong Kong Polytechnic University, as a Research Fellow. From 2001 to 2014, he was with Hebei University as a Professor and the Dean of the School of Mathematics and Computer Sciences. He was the Founding Director of the Key Laboratory on Machine Learning and Computational Intelligence, Hebei. He was a Distinguished Lecturer of the IEEE SMCS. Since 2014, he has been a Professor with the Big Data Institute, Shenzhen University. He has edited over ten special issues and authored or co-authored over three monographs, two textbooks, and over 200 peer-reviewed research papers. As a Principle Investigator (PI) or co-PI, he has completed over 30 research projects. His research interests include uncertainty modeling and machine learning for big data. He is the previous BoG Member of the IEEE SMC Society. He was a recipient of the IEEE SMCS Outstanding Contribution Award in 2004 and the IEEE SMCS Best Associate Editor Award in 2006. He is the Chair of the IEEE SMC Technical Committee on Computational Intelligence and the General Co-Chair of the 2002–2017 International Conferences on Machine Learning and Cybernetics, co-sponsored by the IEEE SMCS. He is the Chief Editor of the *Machine Learning and Cybernetics Journal* and an associate editor of a couple of journals in related areas. He has supervised over 100 M.Phil. and Ph.D. students. According to Google scholar, the total number of citations is over 5000 and the maximum number of citation for a single paper is over 200. He is on the list of Elsevier 2015/2016 most cited Chinese authors.

Chien-Chang Hsu (M'07) received the M.S. and Ph.D. degrees from the National Taiwan University of Science and Technology in 1992 and 2000, respectively. He is currently a Professor with the Department of Computer Science and Information Engineering, Fu Jen Catholic University, Taiwan. He is also the Director of the Information Technology Center. His research interests include machine learning, intelligent systems, medical image processing, and medical informatics. He is the Chair of the Medical Informatics and Innovative Applications Program, Fu Jen Catholic University.

AUTHOR QUERIES

AUTHOR PLEASE ANSWER ALL QUERIES

PLEASE NOTE: We cannot accept new source files as corrections for your paper. If possible, please annotate the PDF proof we have sent you with your corrections and upload it via the Author Gateway. Alternatively, you may send us your corrections in list format. You may also upload revised graphics via the Author Gateway.

AQ:1 = Please provide the postal code for " South China University of Technology, Henan Normal University, Shenzhen University, and Fu Jen Catholic University."
AQ:2 = Please provide the current affiliation for "Daniel S. Yeung."
AQ:3 = Please confirm the volume no. for ref. [20].
AQ:4 = Please confirm the title and also provide the accessed date for refs. [55]–[58].