Face Liveness Detection From a Single Image via Diffusion Speed Model

Wonjun Kim, *Member, IEEE*, Sungjoo Suh, *Member, IEEE*, and Jae-Joon Han, *Member, IEEE*

***Abstract*— Spoofing using photographs or videos is one of the most common methods of attacking face recognition and verification systems. In this paper, we propose a real-time and nonintrusive method based on the diffusion speed of a single image to address this problem. In particular, inspired by the observation that the difference in surface properties between a live face and a fake one is efficiently revealed in the diffusion speed, we exploit antispoofing features by utilizing the total variation flow scheme. More specifically, we propose defining the local patterns of the diffusion speed, the so-called local speed patterns, as our features, which are input into the linear SVM classifier to determine whether the given face is fake or not. One important advantage of the proposed method is that, in contrast to previous approaches, it accurately identifies diverse malicious attacks regardless of the medium of the image, e.g., paper or screen. Moreover, the proposed method does not require any specific user action. Experimental results on various data sets show that the proposed method is effective for face liveness detection as compared with previous approaches proposed in studies in the literature.**

***Index Terms*— Spoofing, diffusion speed, total variation flow, local speed pattern, face liveness detection.**

I. Introduction

With the increasing demand for high-level security in mobile devices, such as smart phones and tablets, biometric techniques have gained considerable attention because of their inherent traits. Thus, iris and fingerprint verification systems have been actively researched [1], [2] and are now deployed in various security systems. Although these approaches guarantee high performance, they require intentional contact with the device, which in the users' experience is inconvenient. As an alternative method, face verification has become the most popular method employed for this task; however, it is vulnerable to diverse spoofing attacks that use photographs or videos of the valid user, which can be easily obtained from the Internet or by capturing him/her using a camera. For example, printed photos, mimic masks, and screenshots, as shown in Fig. 1, are readily employed for malicious log-in attempts. In addition, an attacker can capture video sequences containing involuntary facial gestures, e.g., eye blinking, of the valid user and replay it to penetrate the security system (see Fig. 1(d)). To address this limitation, several researchers have devoted considerable effort to discriminating live faces from fake ones based on motion, spectrum, and image quality information.

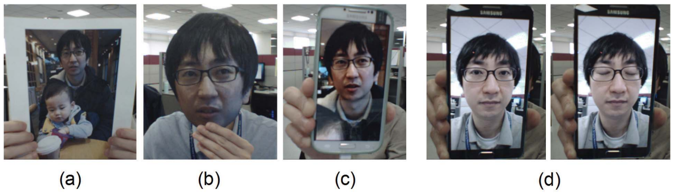


Fig. 1. Examples of images used in spoofing attacks to log-in to mobile devices. (a) Printed photo. (b) Mimic mask. (c) Screenshot. (d) Video replay containing eye blinking.

First, motion-based approaches are most commonly employed for anti-spoofing. They aim at detecting the natural responses of the face, which include eye blinking [3], [4], mouth movement [5], and head rotation [6]. Specifically, Pan *et al.* [3] detected eye blinking based on the undirected conditional graphical framework, in which a discriminative measure of eye states is incorporated. In [5], the authors proposed utilizing the optical flow line of the mouth region. They projected velocity vectors onto their intuitive stick-mouth model and extracted the statistics of the lip motion for face liveness detection. Anjos *et al.* [29] proposed utilizing correlations between the foreground and background regions obtained from the optical flow. Specifically, they attempted to detect motion correlations between the head of the user and the background that indicate a spoofing attack. Although these approaches are conceptually simple, multiple frames are required to track face components, which leads to an increase in the detection time, and highly cooperative user actions are also required. On the other hand, spectrum-based methods clearly expect the inter-class difference between live and fake faces by selecting appropriate working spectrums. In [7], the authors measured the reflectance disparities between live and fake faces based on the computed radiance under different illuminations, and these estimated values were then applied to the Fisher linear discriminant. Zhang *et al.* [8] also measured the albedo curves of different materials, i.e., skin and non-skin, and selected two discriminative wavelengths. As mentioned, these approaches may lead to correct liveness detection; however, they require additional devices (e.g., a near infrared sensor), which are not easily deployed in mobile systems.

Manuscript received August 4, 2014; revised January 14, 2015 and March 26, 2015; accepted April 2, 2015. Date of publication April 13, 2015; date of current version April 29, 2015. The associate editor coordinating the review of this manuscript and approving it for publication was Prof. Richard J. Radke.

The authors are with the Multimedia Processing Laboratory, Samsung Advanced Institute of Technology, Seoul 136-701, Korea (e-mail: wonjun96.kim@samsung.com; sungjoo.suh@samsung.com; jaejoon.han@samsung.com).

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/TIP.2015.2422574

1057-7149 © 2015 IEEE. Personal use is permitted, but republication/redistribution requires IEEE permission.
See http://www.ieee.org/publications_standards/publications/rights/index.html for more information.

Finally, image quality-based approaches assume that fake faces tend to be more seriously distorted by the imaging system and thus yield a lower quality image under the same capturing condition. Li *et al.* [9] proposed using coefficients of Fourier transform based on the observation that fake faces lose more details, i.e., high-frequency components, since the majority are captured twice by the camera. Similarly, the authors of [10] applied multiple DoG filters to extract adequate frequency information to determine the liveness of the given face image. In particular, Tan *et al.* [11] combined texture information with the response of the DoG filter to improve the performance of face liveness detection. Inspired by this work, Peixoto *et al.* [12] applied a similar scheme, defined as a combination of DoG filters and a standard sparse logistic regression model, to bad illumination conditions. Maatta *et al.* [13] attempted to extract micro textures by using the multiscale local binary patterns (LBP), which are frequently treated as a liveness clue. Further, in [14], the authors extracted these micro textures in the regions of face components, e.g., eyes and nose. Recently, Marcel and co-workers used their reliable dataset, which contains various types of spoofing attacks frequently occurring in real-world scenarios, to research anti-spoofing methods. They further provided a novel evaluation framework for a verification system to detect spoofing attacks [27], [28]. Although these methods are effective when a single image is used, they are vulnerable to high resolution-based spoofing attacks that use a large display (see Fig. 1(c) and (d)).

To address the above problems, we propose a novel and simple method for detecting face liveness from a single image. The key idea of the proposed method is that the difference in surface properties between live and fake faces can be efficiently estimated by using diffusion speed. Specifically, we propose computing the diffusion speed by utilizing the total variation (TV) flow scheme and extracting anti-spoofing features based on the local patterns of diffusion speeds, the so-called local speed patterns (LSPs). Our features are subsequently input into a linear SVM classifier to determine the liveness of the given face image. As compared to previous approaches, the proposed method performs well regardless of the image medium and even under varying illuminations. This is quite desirable for achieving robust face recognition and verification in a wide range of environments. The experimental results on various datasets demonstrate that our proposed method provides a reliable performance of face liveness detection. The rest of this paper is organized as follows. The proposed liveness detection scheme with our diffusion speed model is explained in detail in Section II. A description of tests on various datasets to demonstrate the efficiency and robustness of our face liveness detection scheme is given in Section III, and the conclusion follows in Section IV.

II. Face Liveness Detection

A. Motivation

The rationale behind the proposed method is that the illumination characteristics of live and fake faces are significantly different, as shown in Fig. 2. It is easy to see that the light on a live face is quite randomly reflected because of the 3D structures (e.g., nose, lip, etc.), whereas the reflectance of the light on a 2D fake face is relatively uniform. This leads to a difference in the illumination effects of captured images of live and fake faces. In order to estimate this difference in a single image, we propose exploiting the concept of diffusion. This is because the illumination energies on a 2D surface are evenly distributed and thus are expected to diffuse slowly, whereas those on a live face tend to move faster because of their nonuniformity. Therefore, it is considered that the diffusion speed, e.g., the difference in pixel values between the original and diffused images, provides useful clues that can be used to discriminate a live faces from a fake one in a single image. In particular, we attempt to model this diffusion process by allowing for the total variation (TV) flow scheme, and extract anti-spoofing features based on the local patterns of the diffusion speed values computed at each pixel position. In the following, we explain the proposed method in detail.

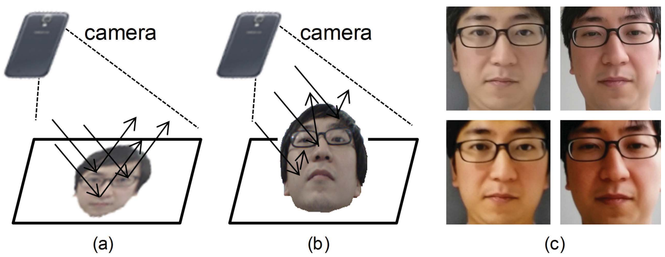


Fig. 2. (a) and (b) Different characteristics of illuminations on a fake and a live face, respectively. (c) Face images captured by a smart phone with full HD resolution (top: live face; bottom: fake face generated by a smart phone screen (i.e., fake faces pass through the camera of the smart phone twice)).

B. Diffusion Speed

In this subsection, we aim to efficiently show the diffusion speed in which illumination characteristics are clearly revealed. To this end, we first conduct nonlinear diffusion on the original face image I, given as [15]:

$$u^{k+1} = u^k + \mathrm{div}(d(|\nabla u^k|\nabla u^k), \quad u(k=0) = I, \tag{1}$$

where k denotes the iteration number. For the diffusivity function $d(\cdot)$, we propose adopting the total variation (TV) flow, defined as [16]

$$d(x) = \frac{1}{x+\xi}, \tag{2}$$

where ξ is a small positive constant. In a given image, the TV flow has been proven to comply with the following rules [17]. 1) Pixels belonging to a small region move faster than those belonging to a large region, e.g., a homogenous region, and 2) the two boundary pixels adapt their value with half that speed. These rules lead to a very useful consequence: by simply computing the difference in pixel values of the original and diffused images generated by the TV flow, we can easily estimate the relative diffusion speed of each pixel.

An important issue is to solve the diffusion equation defined in (1). To this end, we use an efficient approach, called the

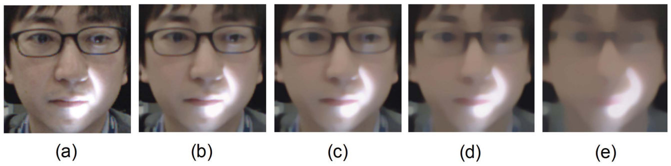


Fig. 3. Examples of diffused image u^k with different iteration numbers. (a) Original image. (b) $k = 2$. (c) $k = 5$. (d) $k = 10$. (e) $k = 20$. Note that illumination effects are efficiently revealed in the diffusion space.

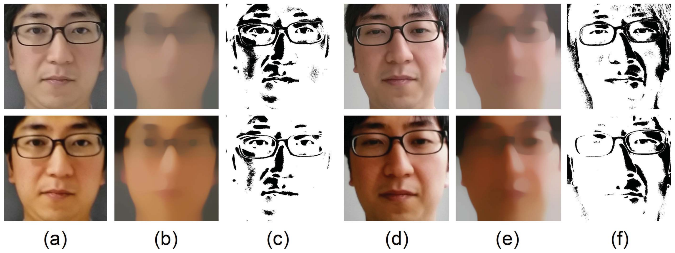


Fig. 4. Diffusion speed maps for live (top) and fake faces (bottom). (a) and (d) Original images. (b) and (e) Diffused image with $L = 10$ and $\tau = 30$. (c) and (f) Binarized diffusion speed values (using the predefined threshold value 0.2 in this case). Note that the black color represents pixels diffusing faster, while the white one represents slower pixels.

additive operator splitting (AOS) scheme [18] defined as

$$u^{k+1} = \frac{1}{2}((I - 2\tau A_x(u^k))^{-1} + (I - 2\tau A_y(u^k))^{-1})u^k, \quad (3)$$

where A_x and A_y denote the diffusion matrices computed in the horizontal and vertical directions, respectively (for more details, see [18]). As compared to the traditional Euler scheme, this AOS scheme is unconditionally stable, and thus, it is possible to use a large time step, e.g., $\tau = 30$, which provides a good compromise between efficiency and accuracy, to enable fast diffusion. Examples of diffused images are shown in Fig. 3. It is noteworthy that our diffusion space successfully reveals the illumination effects (see Fig. 3(d) or (e)).

In the following, we define the diffusion speed at each pixel position (x, y), which represents the amount of difference on the log space between the diffused image and the original one, given

$$s(x, y) = |\log(u^0(x, y) + 1) - \log(u^L(x, y) + 1)|, \quad (4)$$

where L denotes the total number of iterations, which is set experimentally in our implementation (see Section III-B). Although the optimal iteration for L can be adaptively determined, e.g., by utilizing the higher order statistics of the diffusion map [20], [21], we simply fix the iteration number L in this study to achieve fast computation, since the positions of the underlying structures of the face of different individuals are similar. It should be emphasized that our diffusion speed is defined on the log space because of its ability to consistently represent the face under varying lighting conditions [19]. The performance variations according to the L and τ settings are given in Section III. The diffusion speed maps of live and fake faces are shown in Fig. 4. To show the results more clearly, we provide a binarized version of diffusion speed maps according to the predefined value, which was set to 0.2 in this experiment, as shown in Fig. 4(c) and (f). Specifically, we can see the difference in the diffusion speeds of fake and live faces, e.g., in the eye and cheek region, although the corresponding original images appear very similar to each other. In particular, we also compared our scheme with the logarithmic total variational (LTV) framework [22], which is popularly employed for illumination normalization, as shown in Fig. 5. As can be seen, the difference between live and fake faces on the reflectance maps produced by the LTV model is negligible as compared to our results (see Fig. 4(c)), which may lead to a performance degradation in liveness detection (see Table III). More examples are shown in Fig. 6. In this case, we compared live faces with fake ones under various illuminations. The results shown in Fig. 6 confirm that our diffusion speed model successfully captures the subtle difference between high-quality fake and live faces even in diverse lighting conditions. Therefore, it is considered that our diffusion speed model will lead to successful face liveness detection in various environments. It should be noted that the live faces were captured by a smart phone with full HD resolution while the fake ones were generated by capturing the screen of a smart phone and a photograph, i.e., fake faces pass through the imaging system twice, in Figs. 4, 5, and 6.

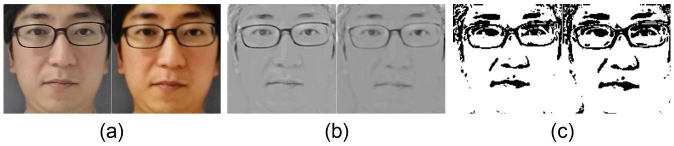


Fig. 5. Reflectance maps for live (left) and fake (right) faces produced by the LTV model [22]. (a) Original face images. (b) Reflectance maps produced by the LTV model [22]. (c) Binarized reflectance maps (using the predefined threshold value 0.1 in this case).

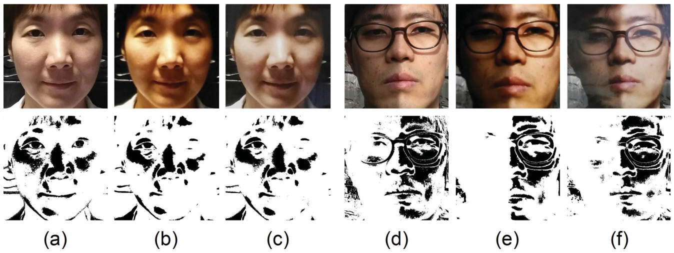


Fig. 6. More examples of the diffusion speed model (top: original images; bottom: binarized diffusion speed maps). (a) and (d) Live faces. (b) and (e) Fake faces captured by the smart phone screen. (c) and (f) Fake photographed faces.

C. Feature Extraction: Local Speed Patterns

On the basis of the above analysis, we can utilize the ability of the diffusion speed model to efficiently extract anti-spoofing features. More specifically, we straightforwardly employ the value of the diffusion speed itself at each pixel position as our baseline features, given as

$$\mathbf{F}_{\text{base}} = \{s(x, y)|0 < x \le W, 0 < y \le H\}, \quad (5)$$

where W and H denote the width and height of the detected face region, respectively. We propose defining the local speed

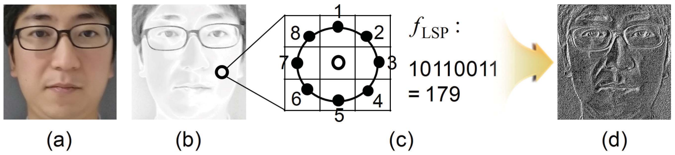


Fig. 7. (a) Original face. (b) Diffusion speed map scaled from [0, 255]. Note that the dark color indicates faster moving pixels. (c) Procedure for computing f_{LSP} at each pixel position. Numbers marked with black dots denote the index of neighborhoods. (d) LSP image.

patterns to efficiently capture even small differences between the diffusion speed maps of live and fake faces as

$$f_{\mathrm{LSP}}(x,y) = \sum_{1\le i\le n} 2^{i-1}\mathrm{LSP}^i(x,y), \tag{6}$$

$$\mathrm{LSP}^i(x,y) = \begin{cases} 1, & \text{if } s(x,y) > s(x_i,y_i), \\ 0, & \text{otherwise,} \end{cases} \tag{7}$$

where n is the number of sampling pixels in the neighborhood of 3×3 pixels, i.e., $n = 8$ in our implementation. (x_i, y_i) denotes the position of the neighborhood pixels centered at (x, y), where $i \in \{1, 2, \cdots, 8\}$. Thus, the range of $f_{\mathrm{LSP}}(x, y)$ is [0, 255] and can be represented as a gray-scale image (LSP image). Figure 7 shows the overall procedure to generate an LSP image using (6).

In the following, we describe the building of the histogram features based on $f_{\mathrm{LSP}}(x, y)$ values of the image block B, the size of which is $M \times M$ pixels. It should be noted that each block is overlapped by its half size (i.e., $M/2$ pixels) in the horizontal and vertical directions, respectively. It should be emphasized that only uniform patterns, which contain at most two transitions between 0 and 1, are utilized, while all the other patterns are accumulated in one additional bin [23]. Therefore, a dimension reduction from 256 to 59 can be achieved for each image block while the discriminative power is retained. Finally, the LSP histograms generated in each block are subsequently normalized by L_2-norm and numerically defined as $\mathbf{h} = (h_1, h_2, \cdots, h_{59})$ with

$$h_k = \frac{N_k}{\sqrt{\sum_{q=1}^{59}(N_q)^2 + \varepsilon}}, \qquad N_k = \sum_{\substack{(x,y)\in B \\ \widetilde{f}_{\mathrm{LSP}}(x,y)\in k}} 1, \tag{8}$$

where B is a set of pixels in the image block, i.e., $M \times M$ pixels, as mentioned. $\widetilde{f}_{\mathrm{LSP}}(x, y)$ denotes the bin of the uniform patterns generated from $f_{\mathrm{LSP}}(x, y)$. By concatenating all the LSP histograms defined in (8), we can represent the face image as a single vector defined as

$$\mathbf{F}_{\mathrm{LSP}} = (\mathbf{h}_1, \mathbf{h}_2, \cdots, \mathbf{h}_K), \tag{9}$$

where K denotes the total number of image blocks in the given face image, and thus, the dimension of the feature vector becomes $59 \times K$. Then, our feature vector $\mathbf{F}_{\mathrm{LSP}}$ is input into the linear SVM classifier for training and tests. The overall procedure of the feature extraction is shown in Fig. 8.

D. Properties of the LSP-Based Feature Vector

We discuss here the advantages of our proposed features for face liveness detection. For each pixel, LSP efficiently encodes not only the illumination characteristics but also the relationships between this information in local regions. The main properties of LSP-based face representation $\mathbf{F}_{\mathrm{LSP}}$ are summarized as follows.

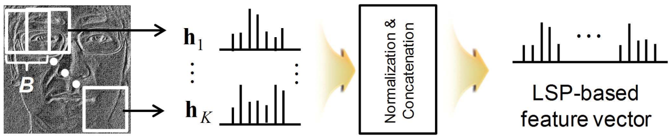


Fig. 8. LSP-based feature vector generation for the given face image. Note that the leftmost illustration shows the LSP image of Fig. 7(a); the dimension of the proposed feature vector is $59 \times K$, where K is the number of image blocks.

1) We focus on the diffusion speed rather than the diffusion result itself, as in the logarithmic total variation (LTV) model [22]. Based on our TV flow-based diffusion speed, which is quite different from the traditional total variational framework used in the LTV model, our method can efficiently reveal the difference in the reflectance characteristics according to the 2D plane and 3D structure, whereas the LTV model provides only the illumination-invariant face image, regardless of the liveness of the given face (see Fig. 5(b)).
2) As compared to the texture patterns widely employed in previous approaches, our LSP-based feature vector captures illumination characteristics on corresponding surfaces. This allows the proposed scheme to be robust to a wide range of spoofing attacks using various media. Moreover, it has a very good ability to discriminate live faces from fake ones, even when the latter are captured in high resolution.
3) Since our diffusion speed model reliably performs under various lighting conditions, as shown in Fig. 6, the LSP-based feature vector can be applied to images in diverse indoor and outdoor environments.
4) Because of the AOS-based diffusion scheme, the proposed method can perform sufficiently well in real-time to be applied in the mobile devices.

These properties allow our LSP-based features to convey reliable information about the given face image to determine whether it is fake or not in real-time on mobile devices.

III. Experimental Results

We first introduce three benchmark datasets: NUAA dataset [11], which is most widely employed in this field; our liveness (SFL) dataset containing real-world scenarios in indoor and outdoor environments with varying illumination conditions; and the Replay-attack dataset [27] composed of photo and videos under different lighting conditions used in attack attempts against 50 clients. Then, we demonstrate the performance of the proposed face liveness detection method on these datasets, with details of parameter choices for our diffusion speed model, e.g., τ (time step) and L (iteration number). Finally, we show examples of our implementation on mobile devices.

A. Datasets

1) NUAA: This dataset [11], which is the most widely adopted benchmark for the evaluation of face liveness

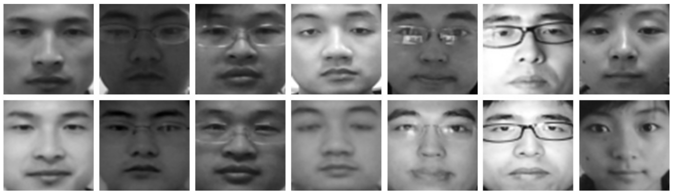


Fig. 9. Samples from the NUAA dataset (top: live faces; bottom: fake faces).

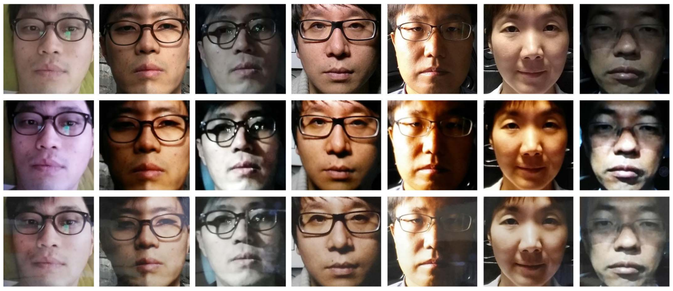


Fig. 10. Samples from the test set of the SFL dataset (top: live faces; middle: the screen of the smart phone; bottom: photographic paper). Note that high-quality fake faces are taken with the full HD camera of a smart phone.

detection, comprises images of 15 subjects who were asked to frontally look at the webcam (capturing faces at 20 fps) with a neutral expression. In addition, none of the faces contains any apparent movement, such as eyeblink or head movement. To create fake examples, the authors of [11] captured pictures of each subject using a usual Cannon camera and printed them on photographic paper and normal A4 paper, respectively. All the faces were detected by using a Viola-Jones detector [24] and geometrically normalized based on the eye localizer [25]. Finally, these images were resized to 64 × 64 pixels with gray-scale representation. Some samples of the NUAA dataset are shown in Fig. 9. For the training set, a total of 3,491 images (live: 1,743 / fake: 1,748) were selected, while the test set was composed of 9,123 images (live: 3,362 / fake: 5,761). It should be noted that there is no overlapping between the training and test sets.

2) SFL: To reflect real-world situations more closely, our dataset was constructed utilizing a smart phone (a Galaxy Note 3) under varying illumination conditions. Specifically, the faces of 21 subjects were captured by a full HD camera mounted on the front side of the smart phone in indoor and outdoor environments. Then, we generated fake faces using two methods, as follows. In the first, we captured the faces shown on the screen of the smart phone using another smart phone. In the second method, we printed the images on a 12.5 cm×17.7 cm sheet of photographic paper and captured an image of the sheet using a smart phone. All the faces were detected by using a Viola-Jones detector [24] and subsequently normalized to 128 × 128 pixels. In the SFL dataset, 608 faces (live: 298 / fake: 310) were collected for training while the test dataset comprised 367 faces (live: 128 / fake: 239). It should be noted that these are mutually exclusive. Samples of our dataset are shown in Fig. 10. As compared to the NUAA dataset, the images in our SFL dataset were captured under more challenging conditions with full HD resolution (i.e., 1920 × 1080 pixels).

TABLE I

NUMBER OF VIDEOS IN THE REPLAY-ATTACK DATASET

Type	Train	Develop	Test	Total
Live	60	60	80	200
Print-attack	30+30	30+30	40+40	100+100
Phone-attack	60+60	60+60	80+80	200+200
Tablet-attack	60+60	60+60	80+80	200+200
Total	360	360	480	1200

Note: numbers marked as summation indicate the amount of hand-held and fixed-support images used in attacks.

TABLE II

PERFORMANCE VARIATION ACCORDING TO PARAMETER SETTINGS ON NUAA AND SFL DATASETS

NUAA	L	τ	Accuracy	L	τ	Accuracy
	5	30	94.76%	10	30	94.53%
	5	40	98.45%	10	35	95.16%
	5	50	94.82%	20	1	90.37%
	5	60	92.57%	20	5	97.74%
	10	10	96.37%	20	10	96.71%
	10	20	97.34%	20	15	78.91%
SFL	L	τ	Accuracy	L	τ	Accuracy
	5	30	82.17%	10	30	89.66%
	5	40	81.28%	10	35	88.94%
	5	50	91.02%	20	10	88.17%
	5	60	88.34%	20	20	89.75%
	10	10	80.43%	20	30	90.63%
	10	20	85.87%	20	40	88.91%

3) Replay-Attack: This dataset consists of 1,300 video clips, the resolution of which is 320 × 240 pixels, of photo and video images under different lighting conditions used in attack attempts against 50 clients. In order to realize the various image media used in attacks, it contains three types of scenarios, i.e., print (printed paper), phone (smart phone screen), and tablet (high-resolution screen) [27]. Specifically, each video was captured in two different environments: fixed-support and hand-held. This dataset is decomposed into three subsets, allowing for training, development, and testing, as shown in Table I, and thus, we can efficiently set the threshold value for the binary classifier. It should be noted that the person identities for each subset do not overlap. As compared to the NUAA and SFL datasets, Replay-attack can be applied to video-based anti-spoofing approaches.

B. Performance Evaluation in NUAA and SFL Datasets

In order to show the performance according to the parameters of our diffusion speed model, we conducted experiments in which the size of the time step and the iteration numbers were varied, as shown in Table II. It should be noted that we use the image block B of 32 × 32 pixels in our implementation and thus the dimension of the feature vector is $59 \times 9 = 531$ and $59 \times 49 = 2,891$ for the NUAA and SFL datasets, respectively. These features are input into the linear SVM classifier [26] for training and testing. In all experiments, we fixed $C = 100$ for SVM, which was shown to give good results when validating the proposed method on a subset of the training set. In Table II, we can see that five iterations are sufficient to yield reliable results in both datasets because of the AOS scheme employed in the proposed method.

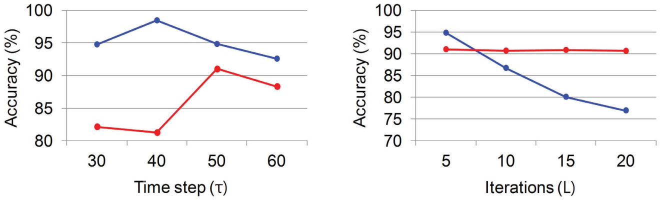


Fig. 11. Performance evaluation when one of the diffusion parameters, i.e., L or τ, is fixed. Left: the number of iterations L is fixed to 5. Right: the time step τ is fixed to 50.

TABLE III
PERFORMANCE COMPARISON OF THE LTV MODEL AND THE PROPOSED METHOD

Methods	LTV [22] (NUAA)	Proposed (NUAA)
Accuracy	68.44%	98.45%
Methods	LTV [22] (SFL)	Proposed (SFL)
Accuracy	74.43%	91.02%

TABLE IV
PERFORMANCE COMPARISON ON THE NUAA DATASET

Methods	DoG-F [9]	DoG-M [10]	DoG-S [11]	DoG-L [12]
Accuracy	84.5%	81.8%	87.5%	94.5%
Methods	MLBP [13]	CP [14]	Baseline	LSP-based
Accuracy	92.7%	97.7%	90.4%	98.5%

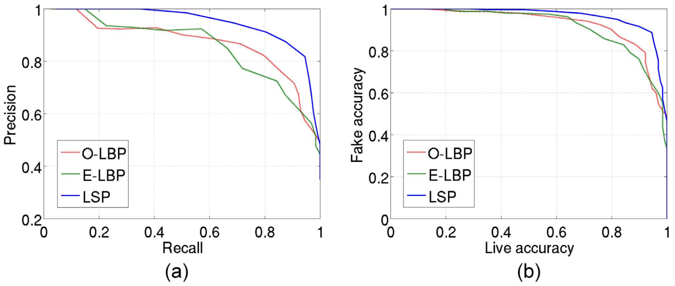


Fig. 12. Performance comparison. (a) Precision and recall curve. (b) Live and fake accuracy curve.

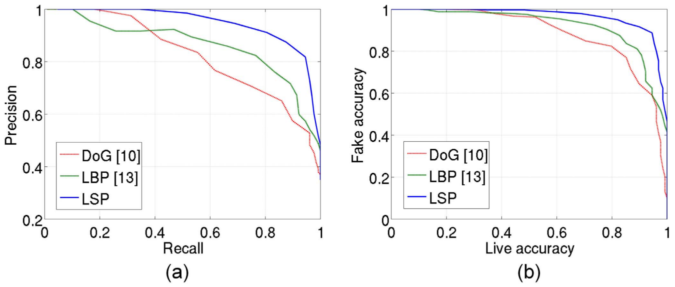


Fig. 13. Performance comparison with DoG-based [10] and LBP-based [13] methods. (a) Precision and recall curve. (b) Live and fake accuracy curve.

It is also noteworthy that a large number of iterations and a large time step do not always guarantee a significant improvement, while they even require more processing times, as shown in Fig. 11. Although the characteristics of these two datasets are quite different from each other, the best performance can be achieved with similar settings, as shown in Table II. Specifically, the detection accuracy is 98.45% when using $L = 5$ and $\tau = 40$ on the NUAA dataset, while $\tau = 50$ with the same number of iterations yields the best performance on our SFL dataset, i.e., 91.02%. Since the positions of the underlying structures in the face image are similar in different individuals, this parameter setting can be applied to other face images without loss of generality. Unless otherwise noted, these values were fixed in our experiments.

To confirm the efficiency and robustness of the proposed method, we first compared it with the LTV diffusion-based scheme [22]. To ensure a fair comparison, we applied the LSP operator to the results of the LTV diffusion to generate the feature vector, which was input into the linear SVM classifier. The corresponding results are shown in Table III. It is easy to see that the proposed approach, i.e., diffusion speed-based feature extraction, performs better than the LTV-based liveness detection. Therefore, we can confirm that our diffusion speed method is highly suitable for capturing the difference between live and fake faces. Moreover, we compared the performance of approach with that of previous approaches on the NUAA dataset: DoG and high frequency-based (DoG-F) [9]; multiscale DoG-based (DoG-M) [10]; DoG and sparse nonlinear regression-based (DoG-S) [11]; DoG and logistic regression-based (DoG-L) [12]; multiscale LBP-based (MLBP) [13]; and component-based (CP) [14]. The performance comparison is shown in Table IV. As can be seen, our LSP-based features achieve the best performance, 98.5%, using only the simple linear classifier. It should be noted that "Baseline" indicates the detection results when using our baseline features $\mathbf{F}_{\text{base}}$ defined in (5); we can see that these also provide quite comparable results. Therefore, it may be considered that the diffusion speed model is very suitable for face liveness detection.

In the following, we describe in more detail our qualitative and quantitative evaluation of the performance of the proposed method using the SFL dataset. To show the advantages of the diffusion speed model in real-world scenarios, we first applied our feature generation scheme explained in Section II-C, which extracts the histogram from overlapped image blocks, to the LBP descriptors of original and edge images. Then, we compared the proposed LSP features with these results using the precision and recall curve, as well as the live and fake accuracy curves, as shown in Fig. 12. It should be noted that the training and tests were conducted using the linear SVM classifier. As can be seen, our diffusion speed model performs better than other low-level descriptors in real-world environments. In particular, the proposed method achieves 91.6% accuracy for fake face detection, while maintaining a 90% detection level for live faces (see Fig. 12(b)). In order to confirm the superiority of the proposed method in real-world scenarios, we also compared ours with two representative approaches driven by DoG [10] and LBP [13], in a manner similar to that shown in Fig. 12 (see Fig. 13). The corresponding detection accuracy is shown in Table V. We can see that the proposed method yields reliable detection results. We also show several examples of images falsely detected by the proposed method in Fig. 14. It is worth noting that a high accuracy for detecting fake faces is preferably required to prevent malicious attacks in face recognition and verification systems. Thus, our method can be successfully deployed in real-world applications.

TABLE V
LIVENESS DETECTION ACCURACY ON THE SFL DATASET

Methods	DoG [10]	LBP [13]	LSP-based
Accuracy	77.5%	87.7%	91.02%
Methods	DoG [10]	LBP [13]	LSP-based
Fake acc.@live acc.=90%	64.4%	81.1%	91.6%

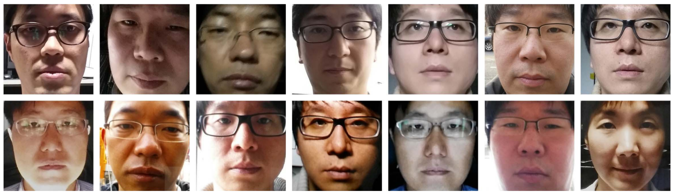


Fig. 14. Top: live faces detected as fake. Bottom: fake faces detected as live.

TABLE VI
CLASSIFICATION ACCURACY OF THE PROPOSED METHOD ON REPLAY-ATTACK

Types	Print-attack	Phone-attack	Tablet-attack
Fixed-support	98.97%	98.89%	85.05%
Hand-held	98.26%	95.23%	89.64%

C. Performance Evaluation in Replay-Attack Dataset

In this subsection, we describe our performance evaluation on the Replay-attack dataset [27], which was designed specifically for face spoofing studies and contains diverse spoofing attacks as well. Unlike other datasets, such as NUAA and SFL, it provides development samples, as well as training and test ones, to efficiently evaluate and compare the performance of anti-spoofing methods. We employed $L = 5$ and $\tau = 40$ for our LSP features yielding the best performance in the NUAA dataset. First, we computed the classification accuracy of the proposed method on the test samples. The corresponding results are shown in Table VI. As can be seen, the performance of the proposed method for detecting spoofing attacks that use printed and digital photographs is very good, while a decrease in the performance occurs for detecting video attacks. In the following, we describe the comparison of our method with previous approaches on this dataset, in which the same protocol employed in [27] and [28] was used. Specifically, we computed the half total error rate (HTER), which is half of the sum of the false rejection rate (FRR) and false acceptance rate (FAR) formulated as

$$\mathrm{HTER}(\tau) = \frac{\mathrm{FAR}(\tau) + \mathrm{FRR}(\tau)}{2}, \qquad (10)$$

where τ denotes the threshold value, which makes the ROC curve. To determine the threshold value for computing HTER values on the test set, we computed the equal error rate (EER), which is defined as a point along the ROC curve where the FAR value equals the FRR value, on the development set illustrated in Fig. 15 and shown in Table VII, respectively, as introduced in [27]. The HTER results based on this value for each spoofing scenario are shown in Table VIII. The HTER values of the proposed method for development and test on the whole set are 13.72% and 12.50%, respectively.

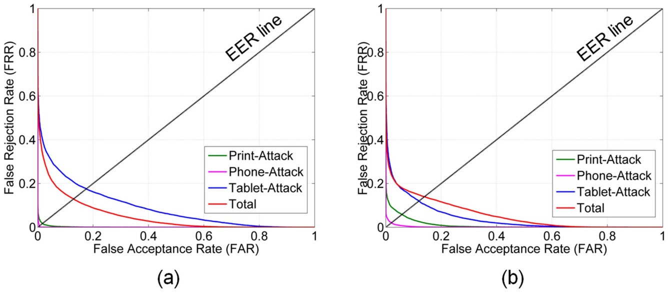


Fig. 15. ROC curves for the Replay-attack dataset. (a) Fixed-support. (b) Hand-held. Note that the equal error rates (EER) are defined at the intersection point between the black line and each ROC curve.

TABLE VII
EER ON THE DEVELOPMENT SET OF THE REPLAY-ATTACK DATASET

Types	Print-attack	Phone-attack	Tablet-attack	Total
Fixed-support	1.79%	0.63%	17.52%	12.88%
Hand-held	5.84%	2.18%	12.09%	13.97%

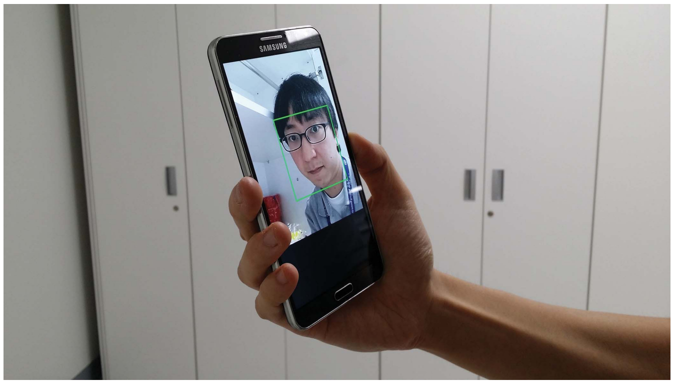


Fig. 16. Demonstration system based on a Galaxy Note 3 smartphone for face liveness detection.

TABLE VIII
HTER OF THE PROPOSED METHOD ON THE REPLAY-ATTACK DATASET

Types	Fixed-Devel.	Fixed-Test	Hand-Devel.	Hand-Test
Print-attack	1.71%	1.70%	5.60%	2.73%
Phone-attack	0.51%	1.31%	1.98%	4.58%
Tablet-attack	16.65%	15.15%	11.63%	11.15%
Total	12.54%	11.62%	11.86%	12.23%

Note: HTER values for development and test on the whole set are 13.72% and 12.50%, respectively.

A performance comparison with previously published results [27] is also shown in Table IX. It should be noted that the previous approaches included in Table IX are explained in detail in [27]. The results shown in Table IX confirm that the proposed method successfully performs under various types of spoofing attacks as compared to previous approaches, which are based on binary patterns of pixel intensities. It is noteworthy that temporal information greatly enhances the performance improvement, as shown in [29], and thus, we consider that a spatiotemporal diffusion scheme would lead to more reliable detection of spoofing attacks, which will be considered in our future work.

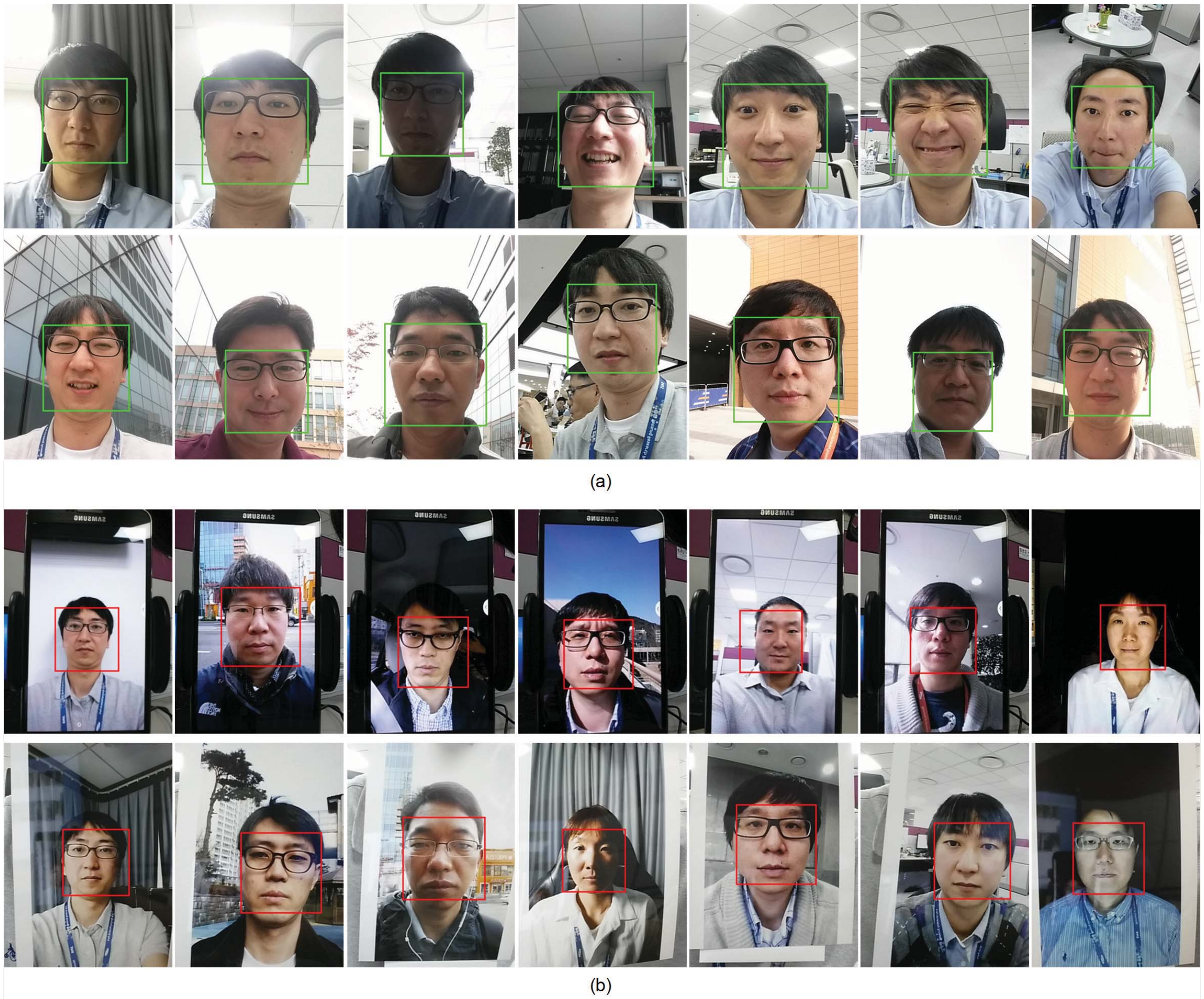


Fig. 17. Demonstration on the smart phone. (a) Detection results for live faces (1st row: variations in poses, illuminations, and expressions; 2nd row: various subjects in indoor and outdoor environments). (b) Detection results for fake faces (3rd row: fake faces from the screen of the smart phone; 4th row: fake faces on photographic paper). Note that the detection results of face liveness are represented by green and red for live and fake faces, respectively. Best viewed in color.

TABLE IX

PERFORMANCE COMPARISON USING THE HTER MEASURE

Methods	Develop	Test
$\mathrm{LBP}^{u2}_{3\times3}+\chi^2$	31.24%	34.01%
$\mathrm{LBP}^{u2}_{3\times3}+\mathrm{LDA}$	19.60%	17.17%
$\mathrm{LBP}^{u2}_{3\times3}+\mathrm{SVM}$	14.84%	15.16%
LBP [13] + SVM	13.90%	13.87%
Proposed method	**13.72%**	**12.50%**

TABLE X

ANALYSIS OF THE PROCESSING TIME FOR THE PROPOSED METHOD

Modules	Diffusion	LSP	Classification	Total
Time	32.7 msec	0.9 msec	0.4 msec	34.0 msec
Ratio	96.2%	2.6%	1.2%	100%

TABLE XI

COMPARISON OF THE PROCESSING TIME

Methods	DoG-based [10]	LBP-based [13]	LSP-based
Time	30.7 msec	33.2 msec	34.0 msec

D. Processing Time

In this subsection, we verify the efficiency of our proposed method. To analyze the processing time in detail, we divided the proposed scheme into three main steps: generation of diffusion speed map, LSP-based feature extraction, and classification. For the classification, we employ the linear SVM classifier, as mentioned. The framework of the proposed method was implemented on a single low-end PC (2.3 GHz CPU and 2.9 GB RAM without parallel processing) with Visual Studio 2010 (C implementation). The average processing time for the test set of the SFL dataset is shown in Table X. These results show that the performance of the proposed method may be considered sufficient to allow its implementation in real-time applications. It should be noted that the diffusion process is the element that consumes most of the processing time. We also show a comparison of the processing time of the proposed method with that of other methods in Table XI. Table XI shows that the difference in the processing time of

TABLE XII
CONFIDENCE VALUES FROM THE SVM CLASSIFIER FOR EXAMPLES SHOWN IN FIG. 17

	1^{st}_{col}	2^{nd}_{col}	3^{rd}_{col}	4^{th}_{col}	5^{th}_{col}	6^{th}_{col}	7^{th}_{col}
1^{st}_{row}	1.53	0.69	1.47	1.28	1.69	1.59	1.62
2^{nd}_{row}	1.11	1.16	0.67	0.54	0.36	1.94	1.35
3^{rd}_{row}	-3.67	-2.26	-3.55	-0.37	-4.39	-2.48	-1.87
4^{th}_{row}	-4.15	-3.39	-1.58	-1.65	-0.41	-1.31	-1.98

Note: the positive values are determined as live while negative ones are regarded as fake by the SVM classifier.

the proposed and other approaches is negligible, while our approach significantly outperforms previous ones. Therefore, we can conclude that our LSP-based face liveness detection can be deployed in various mobile devices.

E. Demonstration on the Mobile Device

In this subsection, we show some examples obtained from the implementation of the proposed method on a Galaxy Note 3 smartphone. The practical demonstration was conducted as shown in Fig. 16. First, we tested our mobile system using images with various person poses and expressions and illuminations, as shown in Fig. 17. Moreover, various subjects, whose images are not included in the training set of the SFL dataset, attempted to verify their liveness in indoor and outdoor environments by using the proposed framework. For testing fake faces, we held faces displayed on a screen and on photographic paper in front of the proposed system, as shown in the last two rows of Fig. 17. It should be noted that green denotes the live face determined by the proposed method, while red indicates the fake face. The confidence values obtained from the linear SVM classifier for each output in Fig. 17 are also shown in Table XII. Our demonstration clearly shows that the LSP-based scheme is quite robust to challenging environments, even under diverse lighting conditions. It should be emphasized again that the proposed method performs in free-view and thus users do not need to take any specific action. This is quite desirable for applying our scheme to real-world scenarios for face recognition and verification in mobile devices.

IV. CONCLUSION

A simple and robust method for face liveness detection was proposed in this paper. The key idea of the proposed method is to adopt diffusion speed for modeling the difference in the illumination characteristics of live and fake faces. Specifically, we proposed exploiting the TV flow and AOS scheme to efficiently compute the diffusion speed, which is robust to varying lighting conditions. To capture the difference between live and fake faces more effectively, we attempted to encode the local pattern of diffusion speed values, the so-called local speed pattern (LSP), and define it as our feature. Based on diverse experimental results, we confirmed that the proposed method successfully performs when the images are captured in a wide range of indoor and outdoor environments, and when they include persons with varying poses and expressions and under different illuminations. Moreover, our LSP-based scheme is effective in real-time and can thus be deployed in various mobile devices. Therefore, we conclude that the proposed method for face liveness detection will lead to high-level security for mobile devices.

ACKNOWLEDGMENT

(Portions of) the research in this paper used the Replay-Attack Corpus made available by the Idiap Research Institute, Martigny, Switzerland.

REFERENCES

[1] A. K. Jain, S. Prabhakar, L. Hong, and S. Pankanti, "Filterbank-based fingerprint matching," *IEEE Trans. Image Process.*, vol. 9, no. 5, pp. 846–859, May 2000.

[2] Y. Wang, J. Hu, and D. Phillips, "A fingerprint orientation model based on 2D Fourier expansion (FOMFE) and its application to singular-point detection and fingerprint indexing," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 29, no. 4, pp. 573–585, Apr. 2007.

[3] 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)*, Oct. 2007, pp. 1–8.

[4] L. Sun, G. Pan, Z. Wu, and S. Lao, "Blinking-based live face detection using conditional random fields," in *Proc. Adv. Biometrics*, Oct. 2007, pp. 252–260.

[5] K. Kollreider, H. Fronthaler, M. I. Faraj, and J. Bigun, "Real-time face detection and motion analysis with application in 'liveness' assessment," *IEEE Trans. Inf. Forensics Security*, vol. 2, no. 3, pp. 548–558, Sep. 2007.

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

[7] Y. Kim, J. Na, S. Yoon, and J. Yi, "Masked fake face detection using radiance measurements," *J. Opt. Soc. Amer. A*, vol. 26, no. 4, pp. 760–766, Apr. 2009.

[8] Z. Zhang, D. Yi, Z. Lei, and S. Z. Li, "Face liveness detection by learning multispectral reflectance distributions," in *Proc. IEEE Int. Conf. Autom. Face Gesture Recognit. (FG)*, Mar. 2011, pp. 436–441.

[9] J. Li, Y. Wang, T. Tan, and A. K. Jain, "Live face detection based on the analysis of Fourier spectra," *Proc. SPIE, Biometric Technol. Human Identificat.*, pp. 296–303, Aug. 2004.

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

[11] 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)*, 2010, pp. 504–517.

[12] B. Peixoto, C. Michelassi, and A. Rocha, "Face liveness detection under bad illumination conditions," in *Proc. 18th IEEE Int. Conf. Image Process. (ICIP)*, Sep. 2011, pp. 3557–3560.

[13] J. Maatta, A. Hadid, and M. Pietikainen, "Face spoofing detection from single images using micro-texture analysis," in *Proc. IEEE Int. Joint Conf. Biometrics (IJCB)*, Oct. 2011, pp. 1–7.

[14] J. Yang, Z. Lei, S. Liao, and S. Z. Li, "Face liveness detection with component dependent descriptor," in *Proc. IEEE Int. Conf. Biometrics (ICB)*, Jun. 2013, pp. 1–6.

[15] P. Perona and J. Malik, "Scale-space and edge detection using anisotropic diffusion," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 12, no. 7, pp. 629–639, Jul. 1990.

[16] M. Rousson, T. Brox, and R. Deriche, "Active unsupervised texture segmentation on a diffusion based feature space," in *Proc. IEEE Comput. Soc. Comput. Vis. Pattern Recognit. (CVPR)*, vol. 2. Jun. 2003, pp. II-699–II-704.

[17] T. Brox and J. Weickert, "A TV flow based local scale measure for texture discrimination," in *Proc. 8th Eur. Conf. Comput. Vis. (ECCV)*, May 2004, pp. 578–590.

[18] J. Weickert, B. M. T. H. Romeny, and M. A. Viergever, "Efficient and reliable schemes for nonlinear diffusion filtering," *IEEE Trans. Image Process.*, vol. 7, no. 3, pp. 398–410, Mar. 1998.

[19] E. H. Land and J. J. McCann, "Lightness and retinex theory," *J. Opt. Soc. Amer.*, vol. 61, no. 1, pp. 1–11, 1971.

[20] W. Kim and C. Kim, "A texture-aware salient edge model for image retargeting," *IEEE Signal Process. Lett.*, vol. 18, no. 11, pp. 631–634, Nov. 2011.

[21] W. Kim and C. Kim, "Active contours driven by the salient edge energy model," *IEEE Trans. Image Process.*, vol. 22, no. 4, pp. 1667–1673, Apr. 2013.

[22] T. Chen, W. Yin, X. S. Zhou, D. Comaniciu, and T. S. Huang, "Total variation models for variable lighting face recognition," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 28, no. 9, pp. 1519–1524, Sep. 2006.

[23] T. Ojala, M. Pietikainen, and T. Maenpaa, "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.

[24] P. Viola and M. J. Jones, "Robust real-time face detection," *Int. J. Comput. Vis.*, vol. 57, no. 2, pp. 137–154, 2004.

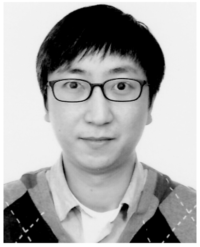
[25] X. Tan, F. Song, Z.-H. Zhou, and S. Chen, "Enhanced pictorial structures for precise eye localization under uncontrolled conditions," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2009, pp. 1621–1628.

[26] C.-C. Chang and C.-J. Lin, "LIBSVM: A library for support vector machines," *ACM Trans. Intell. Syst. Technol.*, vol. 2, no. 3, Apr. 2011, Art. ID 27.

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

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

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



Wonjun Kim (M'13) received the B.S. degree in electrical engineering from Sogang University, Seoul, Korea, the M.S. degree from the Department of Information and Communications, Korea Advanced Institute of Science and Technology (KAIST), Daejeon, Korea, and the Ph.D. degree from the Computational Imaging Laboratory, Department of Electrical Engineering, KAIST, in 2006, 2008, and 2012, respectively.

Since 2012, he has been with the Multimedia Processing Laboratory, Samsung Advanced Institute of Technology, Gyeonggi-do, Korea, where he is currently a Research Staff Member. His research interests include image and video understanding, computer vision, pattern recognition, and biometrics, with an emphasis on saliency detection, face and action recognition. He has served as a regular reviewer for over 20 international journal papers, including the IEEE TRANSACTIONS ON IMAGE PROCESSING, the IEEE TRANSACTIONS ON CIRCUITS AND SYSTEMS FOR VIDEO TECHNOLOGY, and the IEEE TRANSACTIONS ON MULTIMEDIA.

Sungjoo Suh (M'07) received the B.S. and M.S. degrees in electronics engineering from Korea University, Seoul, Korea, and the Ph.D. degree from the School of Electrical and Computer Engineering, Purdue University. Since 2009, he has been a Research Staff Member with the Samsung Advanced Institute of Technology, Suwon, Korea.

His current research interests include image processing, pattern recognition, and biometrics. He has also worked on interactive display architecture, computational photography, analysis for document forensics and printing, and image watermarking. He has also served as a Program Committee Member for VECTaR 2014 in conjuction with European Conference on Computer Vision in 2014.

Jae-Joon Han (M'07) received the B.S. degree in electronic engineering from Yonsei University, Korea, in 1997, the M.S. degree in electrical and computer engineering from the University of Southern California, Los Angeles, in 2001, and the Ph.D. degree in electrical and computer engineering from Purdue University, West Lafayette, IN, USA, in 2006.

He was a Teaching Assistant and then a Research Assistant with the School of Electrical and Computer Engineering, Purdue University, from 2001 to 2006. Since 2006, he has been with Purdue University, where he was a Post-Doctoral Fellow in 2007. He has been with the Samsung Advanced Institute of Technology, Gyeonggi-do, Korea, since 2007, as a Principal Researcher. His research interests include statistical machine learning and data mining, computer vision, and real-time recognition technologies. He also participated in the development of standards, such as ISO/IEC 23005 (MPEG-V) and ISO/IEC 23007 (MPEG-U), and served as the Editor of ISO/IEC 23005-1/4/6.