Copyright ⓒ 2019 KSII

Face Spoofing Attack Detection Using Spatial Frequency and Gradient-Based Descriptor

Zahid Ali [1, 2] and Unsang Park[1*]
[1] Department of Computer Science and Engineering, Sogang University
Seoul, South Korea
[e-mail: {zahid,unsangpark}@sogang.ac.kr]
[2] Department of Electronic Engineering, NED UET
Karachi, Pakistan
[e-mail: szahid@neduet.edu.pk]
*Corresponding author: Unsang Park

Received May 24, 2018; revised August 7, 2018; accepted September 6, 2018;
published February 28, 2019

Abstract

Biometric recognition systems have been widely used for information security. Among the most popular biometric traits, there are fingerprint and face due to their high recognition accuracies. However, the security system that uses face recognition as the login method are vulnerable to face-spoofing attacks, from using printed photo or video of the valid user. In this study, we propose a fast and robust method to detect face-spoofing attacks based on the analysis of spatial frequency differences between the real and fake videos. We found that the effect of a spoofing attack stands out more prominently in certain regions of the 2D Fourier spectra and, therefore, it is adequate to use the information about those regions to classify the input video or image as real or fake. We adopt a divide-conquer-aggregate approach, where we first divide the frequency domain image into local blocks, classify each local block independently, and then aggregate all the classification results by the weighted-sum approach. The effectiveness of the methodology is demonstrated using two different publicly available databases, namely: 1) Replay Attack Database and 2) CASIA-Face Anti-Spoofing Database. Experimental results show that the proposed method provides state-of-the-art performance by processing fewer frames of each video.

Keywords: Video-based face spoofing, Face recognition, Fourier transform, SIFT, spatial frequency

This work was supported by Institute for Information & communications Technology Promotion(IITP) grant funded by the Korea government(MSIT) (2017-0-01772. Development of QA system for video story understanding to pass Video Turing Test) and Information & communications Technology Promotion(IITP) grant funded by the Korea government(MSIT) (2017-0-01781. Data Collection and Automatic Tuning System Development for the Video Understanding).

http://doi.org/10.3837/tiis.2019.02.022 **ISSN : 1976-7277**

1. Introduction

Unlike a password-protected security system, biometric-based information security system use behavior, physical movement, or chemical traits to recognize an authorized person. The most common cues of the biometric system are face, iris, periocular [1], fingerprint, voice, and DNA. Recently, face recognition, in particular has gained wide attention and is being extensively applied in security system [2]. Since the release of Ice Cream Sandwich, the Android OS has come with a built-in face authentication system to unlock the mobile phone [3]. However, the vulnerabilities of the security system, which are based on facial characteristics, have been experimentally shown. The Security and Vulnerability Research Team of the University of Hanoi has demonstrated how to spoof and bypass the practical security system using fake facial images of the authorized users [4]. Therefore, addressing the face spoofing attacks is crucial to enhance the security of the facial biometric system to the level of practical use.

A spoofing attack occurs when a person with no authority tries to masquerade an authorized person by falsifying the data captured by acquisition sensors [5]. In the case of iris-based login system, it is difficult to acquire a high-resolution iris image of an authorized person [6], [7]. Spoofing in fingerprint-based login system is also rather difficult, as it requires obtaining the fingerprint of an authorized person and printing it on a special piece of paper or silicon surface [8]. On the other hand, face recognition-based login system are more user-friendly, but are more susceptible to spoof attacks, as the source data of an authorized person can easily be obtained from social network websites, such as Facebook, Twitter, Instagram, etc., or can be directly taken from the user at a distance [9].

In the context of facial biometrics, with regard to the modalities of the data, spoofing attacks can be classified into the following three categories: (1) photograph, (2) video, and (3) 3D facial mask. If anyone of these types of data is successfully used, the facial biometric system becomes fragile [10]. In order to enhance the security of this system, the biometric community has developed many countermeasures. A typical counter-measure against spoofing is movement detection that aims at detecting physiological signs of life, such as eye blinking, facial expression changes, and mouth movements. Another counter-measure consists of combining facial recognition with other biometrics modalities, such as speech. There are also other approaches to detect face spoofing based on the structure from motion using the depth information [5].

In this paper, we propose a fast and robust method to detect spoofing attacks in face recognition-based login system. The proposed approach analyzes 2D Fourier spectra as the countermeasures of video-based spoofing attacks, namely, the replay-video attack and the print-photo attack. To this end, we use gradient-based descriptor to extract discriminant features from frequency domain spectra of the real and fake videos. Then, these features are used to train multiple Support Vector Machine (SVM) classifiers [49], [50] to distinguish between real and fake videos.

This paper is organized as follows. Section 2 overviews related works on face-spoofing detection methods. Section 3 provides a comprehensive description of the proposed method. Section 4 presents the experimental setup using publicly available databases, outlines the experimental results of the proposed method, and compares them with those available with the state-of-the-art system. Section 5 draws conclusions and discusses further research directions.

2. Related Work

Overall, there are four categories of countermeasures for spoofing attacks: (1) data-driven characterization, (2) user behavior modeling, (3) user interaction, and (4) additional devices for multi-modalities [11]. Our proposed method belongs to the data-driven characterization category, which does not require to process consecutive frames in contrast with (2), no user interaction in contrast with (3), and no additional devices as in (4). In what follows, we review some of the major countermeasures for spoofing attacks.

Data-driven characterization methods are based on the analysis of texture and reflectance properties. For example, Li et al. [12] proposed an anti-spoofing method for photo-based spoofing attacks under the assumptions that the size of a photo is usually smaller than a live face and that facial expressions in photos are static. Under these circumstances, spoofing videos would contain fewer high frequency components than real videos. These characteristics can be captured by analyzing the 2D Fourier spectrum. Furthermore, Tan et al. [13] proposed an anti-spoofing method which considers Lambertian reflectance to distinguish between real accessing videos and spoofing-attack videos. This approach is based on the assumption that surface roughness is different between valid accessing videos and spoofing-attack videos. The authors used the variational retinex-based method and difference-of-Gaussian (DoG) to extract latent reflectance features. The authors reported the results on a publicly available database (NUAA Database) that consists of true accesses and spoofing attacks of 15 subjects in two different qualities, i.e., photo and laser-print. In addition, Peixoto et al. [14] proposed a method by extending the technique proposed by Tan and co-authors [13] to detect video-based spoofing attacks. They used the fact that the brightness of the LCD screen affects the re-captured image from the original image. The method proposed a preprocessing step which consists of adaptive histogram equalization before extracting latent reflectance features for capturing the effect of the brightness of the LCD screen. The experimental results on the NUAA Database showed that the proposed method reduces the classification error by over 50% for the quality of laser-print dataset. Furthermore, Maatta et al. [5] proposed a method based on the micro-texture analysis. In this methods, Local Binary Patterns (LBP) for micro-texture analysis and Support Vector Machine (SVM) classifier are used. In the method proposed by Schwartz et al. [15], several feature descriptors to describe facial information are combined. This method focuses on facial regions and extracted holistic feature descriptors to describe facial information, such as shape, color, and texture. Gragnaniello et al. [47] assess the potential of several descriptors including Weber Local Descriptor (WLD) for the liveness detection task in anti-spoofing systems. Another local descriptor called Weber Local Binary Pattern (WLBP) [53], which combines the discriminability of LBP and WLD is effectively used for blink (liveliness) detection [52]. Sometimes, feature learning methods such as [54-56] are also used along with these feature extraction methods to avoid the curse of dimensionality problem.

A new and more challenging face anti-spoofing database called CASIA Face Anti Spoofing Database (CASIA-FASD) was published by Zhang et al. [16], introducing a new type of spoofing attack named Cut Photo Attack. The authors also presented 6 different protocols (see Section 4 for further detail) and the corresponding baseline results. The proposed method uses multiple Difference of Gaussian filters (DoG) to extract high-frequency information from face images, which is treated as the liveness clue. Galbally et al. [17] proposed 14 image quality features to distinguish between real and fake videos. The image quality measure (IQM) includes pixel difference measures, correlation-based measures, edge-based measures, etc. Once the feature vector is generated, the input image is classified as real or fake using a simple Linear Discriminant Analysis (LDA) classifier [51]. However, the IQM approach largely depends on the quality of the spoof video and hence does not yield good results on various data in CASIA-FASD. Similarly to Galbally et al. [17], Wen et al. [18] proposed the use of four different features (specular reflection, blurriness, chromatic moment, and color diversity) to define the image distortion for face-spoof detection. An ensemble classifier with multiple SVMs trained for different face spoof attacks, such as printed photo and replay video, was used for classification. The authors also collected a new database, MSU-MFSD [18] and published baseline results on this dataset. The proposed technique shows good results under both intra- and inter-database settings.

Contrary to using only gray-scale images [19], Boulkenafet et al. [20] recently proposed including the chrominance information alongside with the luminance information while considering texture analysis for face anti-spoofing. The authors used the joint color-texture information from the luminance and chrominance channels using a color LBP descriptor and presented promising results on CASIA and Replay-Attack databases. Extending this work one step ahead, Boulkenafet et al. [21] conducted extensive experiments on the two datasets and MSU-MFSD. In their experiments, they analyzed two more feature descriptors, Co-Occurrence of Adjacent Local Binary Patterns (CoALBP) and Local Phase Quantization (LPQ), and presented the state-of-the-art results.

Rather than using different texture-based features and evaluating their performances, Yang et al. [22] trained a deep convolutional neural network (CNN) that could learn features of a high discriminative ability in a supervised manner. The authors also demonstrated the positive role of the background region in face anti-spoofing by training the CNN with frames of different spatial scales, including the background region other than the face region. The features extracted from the last fully-connected layer of the CNN were then used to train SVM to classify the video as real or spoof. Xu et al. [23] extended this work by incorporating Long Short-Term Memory (LSTM) units with CNN. Their results show a great improvement on CASIA-FASD as compared to the hand-crafted features. The use of CNN presents an effective measure for anti-spoofing; however, the computational complexity in the training phase incurs extra delays that require special hardware for acceleration.

Regarding the methods based on user behavior modeling, typical anti-spoofing methods consider physiological signs, such as eye blinking or movements of facial parts [10] for detecting photo-based spoofing attacks. For example, using the physiological sign of eye blinking that occurs approximately once every 2-4 seconds, Pan et al. [24]proposed a photo-based anti-spoofing attack method. Specifically, they used an undirected conditional random field framework to model eye-blink with hidden Markov models. Furthermore, Tirunagari et al. [25] used the property of dynamic mode decomposition (DMD) to represent temporal information of entire video as a single image with the same dimension as the images contained in the video for the purpose of capturing liveness cues. Tirunagari et al. claimed that, unlike Principle Component Analysis (PCA), DMD treats a video as a sequence of images and projects them in the principle motion subspaces. As a result, DMD is superior to PCA in classifying motions. The authors proposed a classification pipeline that consists of DMD, LBP feature extraction, and SVM-based classification. Although the DMD-based method showed 0% HTER on Replay-Attack dataset when all frames of the complete video (240 frames) were processed, the HTER increased drastically when fewer than 240 frames were used.

In addition, Pinto et al. [26] proposed to capture noise signatures generated by the replay video to distinguish between spoofing attacks and valid access. To capture the noise signature, the authors used visual rhythm on the 2D Fourier spectrum space and extracted the gray-level co-occurrence matrices (GLCM) from the visual rhythm. They reported perfect classification performance on a subset of Print-Attack Database [9]. However, their method requires several frames to compute the visual rhythm. The authors considered only two regions of interest in the 2D Fourier spectrum image to construct two types of visual rhythms, namely: (i) horizontal and (ii) vertical visual rhythm formed by central horizontal and vertical lines, respectively. Pinto et al. recaptured the print-attack dataset from Print-Attack Database using 6 different monitors for the evaluation of their method. However, this biases their evaluation results with respect to the quality of the used LCD screens.

Contextual information comes into play when the quality of texture is not sufficiently good to distinguish between noises generated by recapturing or the low-quality camera or the recapturing device does not produce visible noisy patterns. In this case, scenic cues can be exploited to determine whether any suspicious object is present in the observed scene [27]. Yang et al. [22] and Xu et al. [23] also proved that the use of the background region has a positive effect on face anti-spoofing. Pan et al. [28] used the background region by proposing the scene context analysis method. They considered not only the facial region but also the background scene that is already known for the security system in their experiments. The authors used cues related to facial information, such as eye blinking, and also analyzed the background scene context. Komulainen et al. [29] proposed to detect the spoofing medium in order to detect a spoofing attack. They made use of the fact that we, humans, rely on contextual information (e.g., the presence of hand holding the screen or photo) to perform spoofing detection. In a similar fashion, their algorithm tries to find the scenic cues, including the presence of an attacker, hand, and misalignments in the upper-body and torso, to detect spoofing attacks. Some other work addressed the multi-modal analysis of the system combining facial analysis with other biometric traits, such as speech [30] and gait. Application specific sensors were also used to acquire multi-spectral [31] or near-infrared images [32].

In this study, we also propose to use 2D Fourier spectrum images; however, rather than using the complete spectrum images, we divide the spectrum image into discrete regions. We compute the SIFT descriptors from the discrete regions and train SVM classifiers. At test time, the regions with high classification accuracies are used, which not only reduces the computational cost but also yields state-of-the-art results using only a small number of frames.

3. Proposed method

One of the key foci of the proposed method is to find noisy patterns (including but not limited to moiré patterns) appearing in the spoofing video. Firstly, these noisy patterns are induced in the spoofing video due to the spatial frequency differences between the display and the acquisition device [33]. Spatial frequency is a characteristic of any structure that is periodic across positions in space. It is a measure of how often sinusoidal components (as determined by the Fourier transform) of the structure repeat per unit of distance. Alternatively, the spatial frequency can be defined as the level of detail present in a stimulus per degree of visual angle. A scene with small details and sharp edges contains more amount of high spatial frequency information than the one composed of large coarse stimuli [16], [34]. For example, **Fig. 1** (a) has much detail (i.e., high spatial frequency information), while the high spatial frequency components have been removed in **Fig. 1** (b), (c), and (d). The same effect of spatial frequency difference is observed in **Fig. 1**(e) and (g), especially in the forehead region (**Fig. 1**(f)). **Fig. 1** (h), (i), and (j) are the mesh plots of frequency domain equivalents of **Fig. 1** (a), (d), and (g) respectively (mesh plots are used for illustrating the distribution of low and high frequency components). It is also evident that the removal of high spatial frequency components from **Fig. 1**(h) (frequency domain equivalent of **Fig. 1** (a)) causes the concentration of low-frequency components in the center region of the 2D frequency plot in **Fig. 1** (i) (frequency domain equivalent of **Fig. 1** (d)) and **Fig. 1** (j) (frequency domain equivalent of **Fig. 1** (j)), respectively.

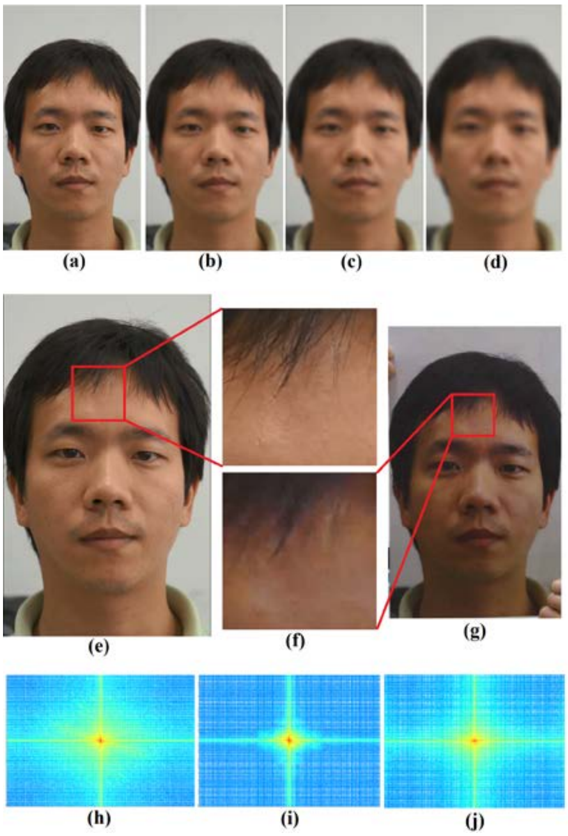

Fig. 1. (a) Original image, (b), (c), and (d) are blurred images of (a). (e) Original video frame, (g) spoof video frame (warped photo attack) from CASIA-FASD and (f) shows a closer look of (e) and (g). (h), (i), and (j) are the frequency domain equivalent images of (a), (d), and (g), respectively.

Secondly, consider the scenario when a camera captures the video of a real-world scene with the face of person ‘A’ (real video). Later, the same camera captures the video (fake video) of the printed photo of person ‘A’ or it recaptures the video of person ‘A’ being played on an LCD (of a cell phone, tablet PC, or desktop monitor). Even though the same camera is used to capture the real video and the fake video, an additional noise will be added while undergoing the capturing process twice [13]. Thirdly, face images printed on the paper surface (whether flat or warped) or the LCD screen are considerably different from that of a live face [12]. These differences also come from the fact that a human face is a 3D object and different parts of the human skin have their own optical qualities (i.e., absorption, reflection, scattering, or refraction), which other materials (e.g., paper, photographic paper, or electronic display) do not possess [6]. According to the Lambertian model [35], the intensity of light reflected from a point on a surface is given by the following (see Eq. (1)):

$$Intensity_{reflected} = L_{illuminated}\left(u_L\right) \cdot \rho \cdot \max\left(u_L \bullet u_I, 0\right) \qquad (1)$$

where $L_{illuminated}$ is the illuminated light intensity coming from the direction u_L, reflected with intensity $Intensity_{reflected}$ from the surface point with albedo ρ (constant of the surface that determines the ratio of light absorbed by the surface) and normal direction u_I. It is clear from Eq. (1) that even if we fix the lighting conditions, the reflected intensity is largely dependent on albedo (ρ) and the surface normal alone. In the case of 3D printed face masks, the normal surface can be made closer to that of a real 3D face, but it will generate a different albedo constant from that the skin of a person. Consequently, the frequency distribution analysis can reveal the difference of the reflectivity of light between the real and the fake video.

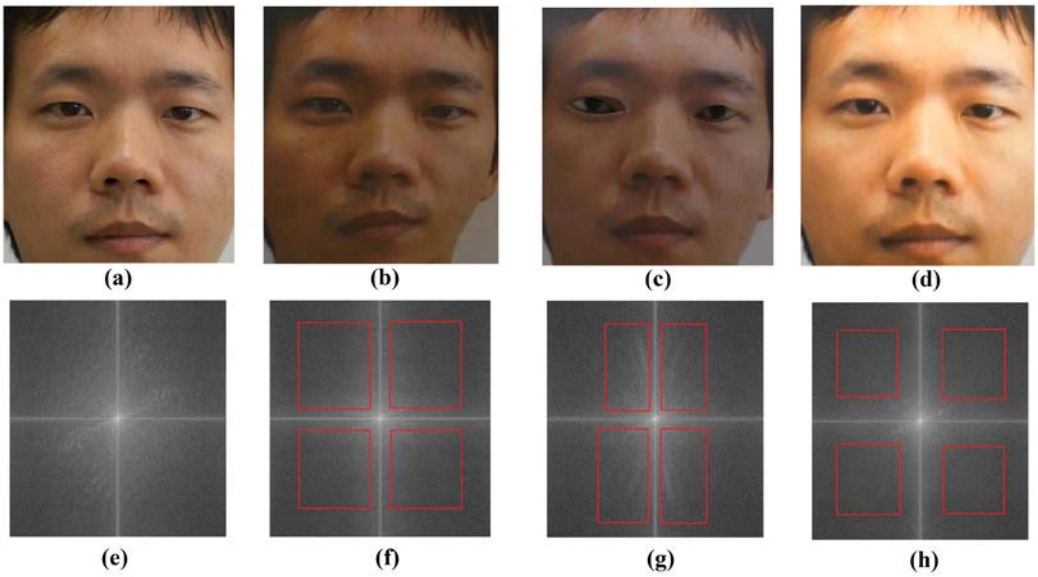

Fig. 2. (a) Original Image (b) Warped Photo Attack (c) Cut Photo Attack and (d) Replay Video Attack. (e), (f), (g), and (h) are the frequency domain images of (a), (b), (c), and (d), respectively. Red boxes in the frequency domain images highlight that the characteristic patterns of spoofing attacks are more prominent in certain regions.

All artifacts or noisy patterns, mentioned above get embedded in the recaptured image as high- and low-frequency spikes at different intervals and, therefore, the Fourier domain analysis can give a closer look at the difference between a real and a fake video. **Fig. 2** shows the difference between a real video frame (**Fig. 2** (a)) and three frames of a spoofing video (**Fig. 2** (b), (c), and (d)). Even though it is difficult to distinguish between a real and a spoofing video in the color domain, their differences become apparent in the frequency domain, as Fourier transform can better handle the small differences in spatial frequencies between the real and the spoof video. Therefore, we use the Fourier transformation to find the noisy patterns and difference of spatial frequency in the spoofing video.

Another novelty of the proposed method is that, rather than processing the complete 2D frequency domain images to classify real and fake video frames, we exploited the fact that the impact of spoofing attacks is more severe in certain regions of 2D frequency domain images. This is evident in the highlighted areas (red boxes) in **Fig. 2** (f), (g), and (h). To detect these specific frequency regions in a frequency image, we used a gradient-based descriptor (i.e., SIFT), as gradient-based descriptors are robust against brightness and small local variations.

This divide-and-conquer approach is widely used in face recognition tasks [48] to alleviate the dilemma of high data dimensionality and small samples.

3.1 Fourier Transformation

The Fourier transformation, which is an extension of the Fourier series, has been widely used in signal processing. A Fourier series is an expansion of a periodic function $f(x)$ in terms of an infinite sum of sin and cosine functions. The Fourier series make use of the orthogonality relationships of the sine and cosine functions. In image processing, the discrete Fourier transform (DFT) converts the pixel values into frequencies, which enables for a separate analysis of low and high frequencies [36]. The low frequency generally represents a brightness of an image and the high frequency represents edges or noise.

For an image of size Height × Width (H × W), the 2-D Discrete Fourier Transform is given by the following (see Eq. (2)):

$$F(x,y)=\sum_{i=0}^{H-1}\sum_{j=0}^{W-1} f(i,j)e^{-l2\pi\left(\frac{xi}{H}+\frac{yj}{W}\right)} \tag{2}$$

where $f(i,j)$ is the image in the color domain and the exponential term is the basis function corresponding to each point $F(x,y)$ in the Fourier space. Eq. (2) represents that the value of each point $F(x,y)$ is obtained by summing up the multiplications of pixel domain values with the corresponding base function. Throughout this paper, we call this $F(x,y)$ the frequency domain image. Also, in our paper, we used the Fast Fourier Transform (FFT) algorithm, which is a faster implementation of the DFT algorithm [36].

3.2 Gradient-based descriptors

Gradient, usually calculated by derivatives, is one of the most fundamental concepts in analyzing images in computer vision. The approach to object detection with gradient descriptors is robust against the changes of brightness, as gradient values are computed from the relative difference with the neighboring pixels.

In our proposed method, we used the descriptors in Scale Invariant Feature Transform (SIFT) to extract features from a 2D Fourier spectrum image. SIFT [37] is a computer vision algorithm to detect and describe local features in images, which is successfully adapted in the object detection and recognition problems. The SIFT algorithm mainly consists of four steps. The first step is scale-space extrema detection from the Difference of Gaussian (DoG) pyramid of the input image. The second step is key-point localization, which includes discarding low contrast and edge response and refining positions of key points. The third step assigns an orientation value to each key-point to achieve invariance to image rotation. The final step assigns a descriptor to each key-point. To create the SIFT keypoint descriptor, a 16×16 neighborhood around the key-point is taken, which is further divided into 16 sub-blocks (4×4 size each). For each sub-block, an 8-bin orientation histogram is created. Therefore, the dimension of the SIFT descriptor vector is 128 bin values.

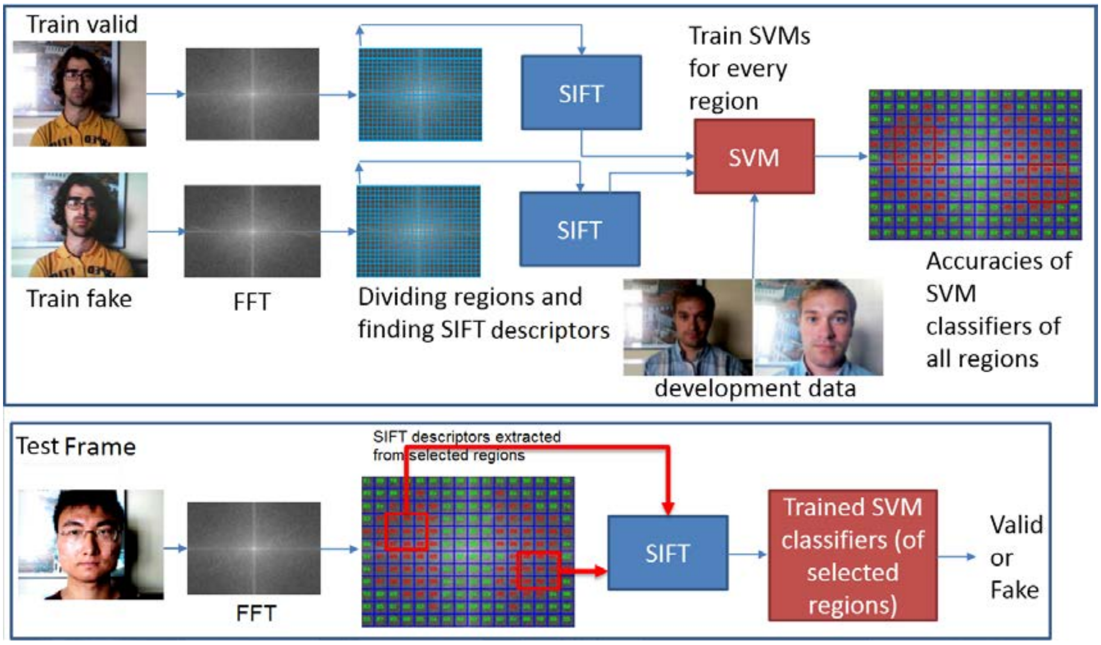
Train valid
Train fake
FFT
Dividing regions and finding SIFT descriptors
SIFT
SIFT
Train SVMs for every region
SVM
Accuracies of SVM classifiers of all regions
development data
Test Frame
FFT
SIFT descriptors extracted from selected regions
SIFT
Trained SVM classifiers (of selected regions)
Valid or Fake

Fig. 3. Overall schematic representation of our proposed method

3.3 Feature extraction and classification method

Fig. 3 shows the overall schematic of the proposed counter-measure. The frames from the train videos were first resized to 320×240 and then converted to frequency domain images. Furthermore, each frequency domain image was divided into equal-sized blocks (i.e., 20×20 pixels), without any overlap area, hence there are 192 blocks in a 320×240 frame. The reason for dividing the frame into smaller blocks is that the noisy patterns are more apparent in small local regions, rather than in the complete frame, in the frequency spectra. Next, we extracted the SIFT descriptors from each block. SIFT extracts 384-dimensional (128×3 channels) feature vectors from each block. Then we trained the SVM classifiers with the RBF kernel for each of the 192 blocks and used the development set to find the accuracies of the SVM classifiers of each block. We extracted 50 frames from each train video for training. The parameters for the RBF kernel were also tunned using the development set. **Fig. 4** shows an example of classification accuracies of the SVM classifiers for all the 192 blocks in a frequency domain image. The accuracies highlighted with red color show the regions where the spatial frequency differences are prominent. Therefore, these discrete regions can effectively discriminate between the fake and the real frames. Those high accuracy regions are concentrated in four areas around the center of the image. According to this result, the center region and corner regions have less effectiveness in detecting the spoofing attacks. Dividing the frequency domain image into smaller blocks and using a small number of blocks with higher classification accuracies enables us to develop a fast and effective face-spoofing detection system.

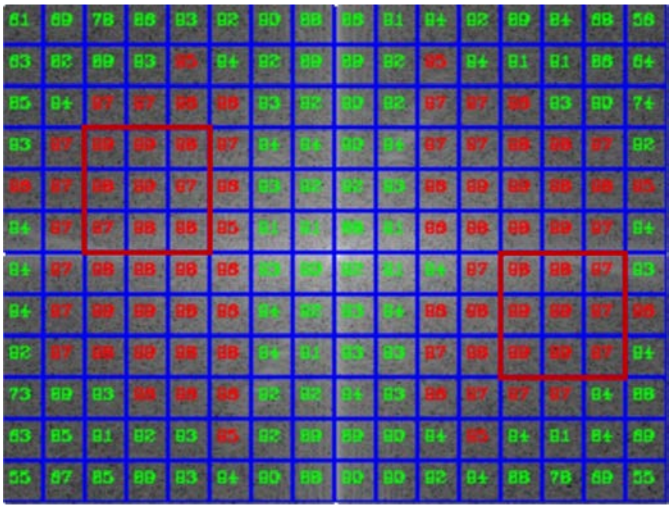

Fig. 4. Classification accuracies of valid and fake videos for each block in a frequency domain image

At the testing step, we extracted the SIFT descriptors only from the blocks that have classification accuracies above the threshold value (= 85%). We calculated the decision score for each selected block region using the trained SVM classifier of that block region. The final decision score for the complete frame was calculated by the sum of the weighted score values from different block regions. Let r_i be the accuracy of each selected box region at training time and $N_{keypoint}$ be the number of selected block regions. The weight value is the probability distribution for the accuracy of the selected block (see Eq. (3)).

$$w_i = \frac{r_i}{\sum_{i}^{N_{keypoint}} r_i} \tag{3}$$

Let s_i be the decision score of each selected box at test time and S_i be the decision score of each frame with $N_{keypoint}$ block regions. Then (see Eq. (4)):

$$S_i = \frac{\sum_{i}^{N_{keypoint}} w_i \times s_i}{N_{keypoint}} \tag{4}$$

Let N_{frame} be the number of frames from a video used for spoofing attack detection, the final decision score S_{final} can be calculated as follows (see Eq. (5)):

$$S_{final} = \frac{\sum_{i}^{N_{keypoint}} S_i}{N_{frame}} \tag{5}$$

In summary, Eq.(4) and (5) correspond to the decision scores on the frame level and the video level, respectively. In the remainder of this paper, we will use the video level decision scores to evaluate the proposed method on the video level.

4. Experimental results

4.1 Databases

We used four publicly available video anti-spoofing datasets to evaluate the proposed method, i.e., CASIA-FASD, Replay Attack DB, 3D Mask Attack Dataset (3DMAD) [43] and Unicamp Video Attack Database (UVAD) [44]. These databases have the characteristics that are mutually exclusive to each other (some of the subjects in Replay Attack DB overlap with those in 3DMAD dataset, but the attack videos in these two datasets completely differ), which helps to develop a generalized video anti-spoofing system to deal with real-world scenarios. We provide a brief introduction to these databases in the following section.

4.1.1 CASIA Face Anti-spoofing DB

This open-source dataset, collected by the Chinese Academy of Sciences (CASIA) Center of Biometrics and Security Research (CASIA-CBSR), contains a total of 50 subjects [16]. The 50 subjects are divided into two subsets: 20 for train, and 30 for test. The real video of each subject is captured in three different qualities, described as High Quality (HQ), Normal Quality (NQ), and Low Quality (LQ). Furthermore, three different types of spoofing attacks (i.e., Cut Photo (CP), Warped Photo (WP) and Replay-Video (RV)) are recorded in three different qualities. Consequently, there are nine fake and three real videos per subject ((9 fake + 3 real) × 50 subjects = 600 videos in total). The CP attack is implemented by capturing the video of the printed photo of a subject, whose eyes portion from the printed photo is cut and the attacker put his/her eyes behind the eyeholes of the printed photo to mimic the eye blinking effect. The WP attack is performed by capturing the video of the printed photo of a subject; the photo is slightly warped, rotated, and moved backward and forward to forge the liveliness effect. The RV attack is performed by recapturing the video of a subject played on an iPad.

4.1.2 Replay Attack DB

This publicly available dataset is provided by the IDIAP Research Institute, which is divided into train, development, and test subsets [9]. The database contains real and fake videos of 50 subjects in two different background conditions: controlled (with uniform background and artificial lighting) and adverse (with non-uniform background and natural illumination). Three different types of attacks are performed, namely: i) mobile attacks are performed by capturing the print photo and video of a subject using iPhone and replaying them on the iPhone screen; ii) print attacks are implemented by videotaping the printed photo of a subject; and iii) high-definition video attacks are performed by displaying the video and photo of a subject on iPad at resolution (1024×768) and recapturing. Based on whether the attacking medium is held by hand or fixed with some rigid support, the fake videos are further divided into i) hand and ii) fixed. Hence, there are a total of 1,000 fake videos (50 subjects × 2 backgrounds × 2 (fixed or hand-held) × (2 mobile attacks + 1 print attack + 2 high definition attacks)) and 200 real videos (50 subjects × 2 backgrounds × 2 sessions).

4.1.3 3D Mask Attack Dataset

This dataset is also publicly provided by IDIAP Research Institute, and similar to Replay Attack DB, this dataset is divided into train, development and test sets [43]. The overall dataset was acquired using a Kinect camera in 3 sessions, out of which first two sessions capture 5 real videos of each of the 17 subjects with frontal-view and neutral expressions. A time delay of 2 weeks is kept between the acquisition of the first two sessions. The third session is dedicated to capturing 5 3D mask attack videos of each of the 17 subjects using the same acquisition device. The protocol to report the results divides the 17 subjects into 3 randomly chosen non-overlapping sets, i.e., the first 7 subjects for training, next 5 subjects for development and last 5 subjects for testing. In total there are 76,500 frames in the complete dataset (3 sessions × 17 subjects × 5 videos per subject × 300 frames per video), out of which 21,000 frames belong to train valid and 10,500 frames belong to train fake set, 15,000 frames belong to development valid and 7,500 frames belong to development fake set, 15,000 frames belong to test valid and 7,500 frames belong to test fake set. Each frame consists of an RGB and a depth image. However, we only used RGB images in our experiments.

4.1.4 Unicamp Video Attack DB

This open dataset consisting of 404 subjects is published by the Institute of Computing at University of Campinas (Unicamp) [44]. A total of 808 valid access videos of 404 subjects are acquired (2 videos per subject with different background) using 6 different cameras, and a total of 16,268 attack videos are acquired using 6 different cameras. The fake videos are displayed on 7 different display devices and each video is recaptured using 6 different quality cameras. The release version of the dataset contains 404 valid access and 9,882 attack videos. We followed the protocol III defined in [45] to report the results. Protocol III divides the dataset into training videos acquired using Sony, Kodak and Olympus cameras, and test videos captured using the other three cameras, i.e., Nikon, Canon and Panasonic.

4.2 Evaluation Results

4.2.1 Evaluation on CASIA Face Anti-spoofing DB

As CASIA-FASD lacks a pre-defined validation set, we have used two-fold cross-validation to train and tune the parameters. We used the first 50 frames of the 10 train videos of the first 10 subjects to train and tune our SVM classifiers by validating them on the rest of the 10 train videos of the next 10 subjects in one fold and interchanged the train and validation sets in the other fold. We also computed the per block accuracy of the SVM classifiers in each fold and averaged them over two folds. Then, the individual results on each CASIA-FASD protocol, as well as overall results on the entire database, were evaluated to compare our results with those available using the state-of-the-art techniques. Individual results on CASIA-FASD protocols are reported by [16], [17], [18], and [20]. Therefore, we summarize our experimental results on CASIA-FASD and the comparison of our method with [16], [17], [18], and [20] in terms of Equal Error Rate (EER) in **Table 1**. We used the maximum voting method to obtain the performances in the last column in **Table 1** as [18]. Buolkenafet et al. [20] used feature averaging, while [16] and [17] did not specify which method they used to obtain the performances for all modalities. The results in **Table 1** show that our proposed approach outperforms the results available with the use of the methods reported in [16], [17], [18] and [20]. It is also notable that our method achieves the performance given in **Table 1** by processing only 10 frames, while [16], [17], and [18] processed 30 or more frames to obtain their results. The results in **Table 1** demonstrate that the proposed method is effective on different levels of certainty, but provides the best performance for each type of attacks among similar studies.

Table 1. Spoofing attack detection performance on CASIA-FASD in EER (%) by the proposed method for different protocols and its comparison with [16], [17], [18], and [20]

Method	Protocol	Warped	Cut	Video	All modalities
Zhang et al. [16]	LQ	-	-	-	13.00
	NQ	-	-	-	13.00
	HQ	-	-	-	26.00
	All Qualities	16.00	6.00	24.00	17.00
Galbally et al. [17]	LQ	25.00	23.30	21.70	31.70
	NQ	23.30	16.70	23.30	22.20
	HQ	10.00	11.70	6.70	5.60
	All Qualities	26.10	18.30	34.40	32.40
Wen et al. [18]	HQ	-	-	-	6.70 (max vote)
Buolkenafet et al. [20]	LQ	-	-	-	7.80
	NQ	-	-	-	10.10
	HQ	-	-	-	6.40
	All Qualities	7.5	5.40	8.10	6.20 (feature averaging)
Proposed	LQ	**5.00**	**3.00**	**6.67**	**3.33**
	NQ	**10.00**	**8.33**	**1.67**	**5.56**
	HQ	**7.25**	**3.64**	**1.82**	**5.56**
	All Qualities	**6.00**	**2.20**	**0.00**	**0.00 (max vote)**

4.2.2 Evaluation on Replay Attack DB

In our experiments on Replay-Attack DB, we also used 50 frames from each video in the train-valid and train-fake dataset to train the SVM classifiers. As the development dataset is separately provided with Replay-Attack database, we used this development dataset to tune the parameters of the RBF kernels and evaluate the classification accuracies of the SVMs of all the blocks per frame. We also set the threshold value at the minimum EER using train and development sets, to find the HTER value from the test set of Replay-Attack DB. The overall experimental results on Replay-Attack DB are shown in **Table 2**. Following the official Replay-Attack protocol, we have presented the results in terms of EER on the development set and Half Total Error Rate (HTER) on the test set, where HTER is given by the following (see Eq. (6)):

$$HTER = \frac{FAR(\tau, D) + FRR(\tau, D)}{2} \tag{6}$$

where $FAR(\tau,D)$ and $FRR(\tau,D)$ represent the false acceptance rate and false rejection rate, respectively, at a certain threshold value τ, on dataset D [19]. The value of τ is estimated on the EER using the development set.

4.2.3 Evaluation on 3DMAD

For the 3D Mask Attack Dataset, there are separate training, development and testing sets, and hence we trained our model using the train set, and fine-tuned our models using the development set. Using the threshold obtained from the development set, we computed the HTER on the test set. We used the first 50 frames from all videos for training purpose, and we achieved best results when we processed 60 frames of the videos in development and test sets. Our spoofing detection scheme achieved HTER of 2.44% on the 3DMAD dataset, which is very close to the state-of-the-art [42] as shown in **Table 2**.

4.2.4 Evaluation on UVAD

Due to the unavailability of separate development set in UVAD, we divided the given train set into a new train set (approx. 70% of the given train set) and a development set (approx. 30% of the given train set). According to Protocol III, the training videos (valid access and attack) from Sony, Kodac and Olympus cameras consist of 344 valid access and 3,528 attack videos. The total test videos from Canon, Nikon and Panasonic cameras comprises of 60 valid access and 6,354 attack videos. We further divide the total train videos into a train set of 240 valid access and 2,470 attack videos, and a development set of 140 valid access and 1,058 attack videos, while we keep the test set the same as defined by Protocol III in [45]. We trained the proposed classifiers using a new train set and fine-tuned using the new development set. Finally, we computed the value of EER on the test set and reported the results in **Table 2**.

4.5 Overall results

The comparison between the proposed method and the state-of-the-art methods on the four datasets are summarized in **Table 2**. As can be seen in the results, the proposed method shows the best performance among similar studies on CASIA FASD and Replay-Attack DB. It is noticeable that the proposed method achieves the results on CASIA FASD given in **Table 2** by processing only 10 frames of the video, while [21], [20], [18], [16], [23], [22],[19], [3], [25], and [17] use 30 or more frames to achieve the results given in **Table 2**.

Table 2. Comparisons of spoofing attack detection performance between the proposed method and state-of-the-art methods on CASIA-FASD, Replay-Attack, 3DMAD and UVAD DB

Method	**Replay-Attack**		**CASIA**	**3DMAD**	**UVAD**
	EER (%)	*HTER (%)*	*EER (%)*	*HTER (%)*	*EER (%)*
CoALBP+LPQ+HSV+YCbCr [21]	**0.00**	3.50	3.20	-	-
LBP+YCbCr+HSV [20]	0.40	2.90	6.20	-	-
LBP+SVM [18]	-	7.41	-	-	-
WLD [47]	-	17.5	-	5.2	
DMD [25]	-	**0.00**	21.75	-	-
LSTM-CNN [23]	-	-	5.17	-	-
DIP [38]	-	5.00	5.07	-	-
CNN [22]	2.14	6.10	4.87	-	-
LBP-TOP [19]	7.90	7.60	10.00	-	-
14-IQF [17]	-	15.20	32.40	-	-
Fisherface [39]	-	-	11.80	-	-
Correlation [3]	-	11.80	30.33	-	-
DoG+SVM [16]	-	-	17.00	-	-
Deep Repesentation [42]		0.75	-	**0.00**	-
Codebook [45]	-	2.75	14.00	8.00	29.87
Deep dictionary via greedy learning (DDGL) [46]	-	0.00	1.3	**0.00**	**16.50**
Proposed method	**0.00**	**0.00**	**0.00**	2.44	26.00

We did not use the depth information for detecting spoofing attack in 3DMAD dataset because our proposed method is based on the analysis of 2D frequency spectrum. Still, our scheme performs very well on the 3D mask attack database as compared to [45], and is closer to the state-of-the-art methods [42][46]. On UVAD dataset, our scheme achieved 26% EER on the test set, which is better than [45]. We believe that the reason for high EER on UVAD dataset is small number of valid videos for training (as small as 20 valid access videos acquired using one type of camera), highly unbalanced numbers of valid and attack videos (20 valid videos captured using Kodac camera against 2,828 attack videos recaptured using Nikon camera), and high false positive in face detections. Thus, as can be seen from **Table 1** and **Table 2**, our proposed method outperforms the state-of-the-art algorithms on CASIA-FAS and Replay Attack databases and shows comparable performances on 3DMAD and UVAD. We will consider using the depth information in our future work on the 3DMAD dataset. For the future work on UVAD dataset, we will employ a better face detection approach instead of using the face locations provided by the dataset, and perform extensive experiments by covering all the protocols of the dataset to handle the multi-modalities of the UVAD dataset.

4.5 Processing overhead

One of the key achievements of our proposed scheme is that it can process the frames at testing time very quickly. At testing time, FFT is applied to the input frame and then the SIFT descriptors are extracted from a small local region in the frequency image. On a core i5 2.40 GHz processor and 8GB RAM, using opencv 2.4.9 implementation of FFT and SIFT and SVM implementation of LibSVM [40], the processing speed is found to be 0.024 seconds/frame (the average of 50 frames of 20 videos). This processing time does not include face detection time, as [18], [41] and [46] in **Table 3** do not include face detection time either. The processing time of most of the published studies on face anti-spoofing is unknown and, therefore, we compare the computational time with only four other studies in **Table 3**. On the other hand, computation of the SIFT descriptors can be parallelized; thus, the processing speed is expected to get faster on a parallel processing platform like FPGA. Despite this, the processing speed of 0.024 seconds/frame is sufficiently fast for real-time processing. It is worth noticing that the proposed method shows worse EER as compared with that of the state-of-the-art method [46] by 9.5%, it is 14.5 times faster than [46], and 1.87 times faster than [42].

Table 3. Comparison of computational cost

Method	Processing Time (sec)	Hardware	Language
Proposed	**0.024**	Core i5 @ 2.40 GHz, 8GB RAM	C++
IDA [18]	0.260	Core i7 @ 2.40 GHz, 8GB RAM	Matlab
DDGL [46]	0.347	Xeon E5-2695 (12 Cores) @ 2.40 GHz, 128GB RAM	Matlab
Deep Representation [42]	0.045	Intel i7 @ 3.50 GHz, Tesla K40 GPU	Mixed implementation (Python, C++, Cuda)
HOOF [41]	0.900	-	Matlab

4.6 Deep learning based methods vs proposed method

If we compare the results in Table 2 and Table 3 simultaneously, we can conclude that our scheme has several benefits over the deep learning based methods:

a) The proposed scheme is lightweight/computationally less expensive as compared to the deep learning based techniques.
b) The proposed scheme does not require special acceleration hardware to enable real-time processing, while the deep learning based methods require acceleration hardware for real-time processing. Table 3 shows that the proposed method achieves 24millisecond processing time with just 4 CPU cores @ 2.4GHz and 8GM RAM, while the deep learning based method [46] achieve 347millisecond processing time even with 12 CPU cores @ 2.4GHz and 128GB of RAM. The scheme in [42] uses Intel i7 @ 3.5GHz with a Tesla K40 GPU.
c) As shown in Table 3, the proposed scheme is 14.5 times and 1.875 times faster than the deep learning methods of [46] and [42], respectively.
d) The deep learning methods typically require multiple hours to complete the training process, while the proposed scheme takes an average of 12 minutes to train the SVM classifiers. The deep learning method of [42] takes about one day to train the deep network, on average.

5. Conclusions

In this study, we proposed a face anti-spoofing method for the three types of face-spoofing attacks, namely i) Replay video attack, ii) Warped photo attack and iii) Cut photo attack. The proposed method uses the spatial frequency of the input frame to classify it as real or fake. We showed that a recaptured photo or a video contains distinctive characteristics due to the double capturing processes, which can be effectively detected by the 2D Fourier spectral analysis. We extracted the SIFT descriptors and trained the RBF kernel-based SVM from different regions of the 2D Fourier spectra; then, only those regions which have high distinctiveness were selected. At the testing time, we extracted the SIFT descriptors from the selected regions only and computed decision scores from them. We used weighted sum of the per-block decision scores for the frame level classification and the sum of the per-frame decision scores for the video level classification. The proposed method achieves state-of-the-art performances on the two publicly available challenging face anti-spoofing databases (0% EER on CASIA-FASD and 0% HTER on Replay-Attack DB) by processing only 10 frames in CASIA, as well as 50 frames in Replay-Attack DB from the input video with minimal processing time (0.024 seconds for each frame). The proposed method also performed reasonably well on 3D Mask Attack and UVAD datasets with unbalanced valid and attack videos and erroneous face detections. Directions for further research include the following: i) testing the proposed method under the cross-database scenario, ii) evaluation on MSU-MFSD and iii) fusion of different gradient-based descriptors.

References

[1] Z. Ali, U. Park, J. Nang, J-S. Park, T. Hong and S. Park, "Periocular recognition using uMLBP and attribute features," *KSII Transactions on Internet and Information Systems*, vol. 11, no. 12, pp. 6133-6151, 2017. Article (CrossRef Link)

[2] A. K. Jain and A. A. Ross, "Introduction to biometrics," *Handbook of Biometrics*, 1st ed., Springer, US, Ch. 1, pp. 1-22, 2008. Article (CrossRef Link)

[3] T. F. Pereira, A. Anjos, J. de Martino and S. Marcel, "Can face anti-spoofing countermeasures work in a real world scenario?" in *Proc. of IEEE Int. Conf. on Biometrics*, pp. 1-8, 2013. Article (CrossRef Link)

[4] M. D. Nguyen and Q. M. Bui, "Your face is not your password: Face authentication bypassing Lenovo – Asus - Toshiba," in *Proc. of Black Hat conference*, pp. 1-16, 2009. Article (CrossRef Link)

[5] J. Maatta, A. Hadid and M. Pietikainen, "Face spoofing detection from single images using micro-texture analysis," in *Proc. of IEEE Int. Joint Conf. on Biometrics*, pp. 1-7, 2011. Article (CrossRef Link).

[6] I. Chingovska, A. Anjos and S. Marcel, "On the effectiveness of local binary patterns in face anti-spoofing," in *Proc. of IEEE Int. Conf. on Biometrics Special Interest Group*, pp. 1-7, 2012. Article (CrossRef Link)

[7] V. R-Albacete, P. T-Gonzalez, F. A-Fernandez, "Direct attacks using fake images in iris verification," *Biometrics and Identity Management*, vol. 5372, pp. 181-190, 2008. Article (CrossRef Link)

[8] T. Matsumoto, H. Matsumoto, K. Yamada and S. Hoshino, "Impact of artificial gummy fingers on fingerprint systems," in *Proc. of SPIE Optical Security and Counterfeit Deterrence Techniques IV*, pp. 1-15, April, 2002. Article (CrossRef Link).

[9] A. Anjos and S. Marcel, "Counter-measures to photo attacks in face recognition: A public database and a baseline," in *Proc. of IEEE Int. Joint Conf. on Biometrics*, pp. 1-7, 2011. Article (CrossRef Link).

[10] G. Pan, Z. Wu and L. Sun, "Liveness detection for face recognition," *Recent Advances in Face Recognition*, InTech, Ch. 9, pp. 235-252, 2008. Article (CrossRef Link)

[11] K. A. Nixon, A. Aimale and R. K. Rowe, "Spoof detection schemes," in *Handbook of Biometrics*, 1st ed., Springer, US, Ch. 10, pp. 403-423, 2008. Article (CrossRef Link)

[12] J. Li, Y. Wang, Y. Tan and A. K. Jain, "Live face detection based on the analysis of Fourier spectra," in *Proc. of SPIE on Biometric Technology for Human Identification*, pp. 296-303, 2004. Article (CrossRef Link).

[13] 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. of European Conference on Computer Vision (ECCV)*, Springer, pp. 504-517, 2010. Article (CrossRef Link)

[14] B. Peixoto, C. Michelassi and A. Rocha, "Face liveness detection under bad Illumination conditions," in *Proc. of IEEE International Conference on Image Processing (ICIP)*, pp. 3557-3560, 2011. Article (CrossRef Link).

[15] W. R. Schwartz, A. Rocha and H. Pedrini, "Face spoofing detection through partial least squares and low-level descriptors," in *Proc. of IEEE International Joint Conference on Biometrics (IJCB)*, pp. 1-8, 2011. Article (CrossRef Link).

[16] Z. Zhang et al., "A face antispoofing database with diverse attacks," in *Proc. of IEEE Int. Conf. on Biometrics (ICB)*, pp. 26-31, 2012. Article (CrossRef Link)

[17] J. Galbally and S. Marcel, "Face anti-spoofing based on general image quality assessment," in *Proc. of IEEE 22nd International Conf. on Pattern Recognition (ICPR)*, pp. 1173-1178, 2014. Article (CrossRef Link).

[18] D. Wen, H. Han and A. K. Jain, "Face spoof detection with image distortion analysis," *IEEE T INF FOREN SEC*, vol. 10, no. 4, pp. 746-761, 2015. Article (CrossRef Link).

[19] T. Pereira, J. Komulainen, A. Anjos, J. M. D. Martino, A. Hadid, M. Pietikainen and S. Marcel, "Face liveness detection using dynamic texture," *EURASIP Journal on Image and Video Processing*, vol. 2014, no. 2, pp. 1-15, 2014. Article (CrossRef Link).

[20] Z. Boulkenafet, J. Komulainen and A. Hadid, "Face anti-spoofing based on color texture analysis," in *Proc. of IEEE Int. Conf. on Image Processing*, pp. 2636-2640, 2015. Article (CrossRef Link)

[21] Z. Boulkenafet, J. Komulainen and A. Hadid, "Face spoofing detection using color texture analysis," *IEEE T INF FOREN SEC*, vol. 11, no. 8, pp. 1818-1830, 2016. Article (CrossRef Link).

[22] J. Yang, Z. Lei and S. Z. Li, "Learn CNN for face anti-spoofing," 2014. Article (CrossRef Link).

[23] Z. Xu, S. Li and W. Deng, “Learning temporal features using LSTM-CNN architecture for face anti-spoofing,” in *Proc. of IEEE 3rd Int. Asian Conf. on Pattern Recognition (ACPR)*, pp. 141-145, 2015. Article (CrossRef Link).

[24] G. Pan, L. Sun, Z. Wu and Y. Wang, “Monocular camera-based face liveness detection by combining eyeblink and scene context,” *Telecommunication Systems*, vol. 47, pp. 215-225, 2011. Article (CrossRef Link).

[25] S. Tirunagari et al., “Detection of face spoofing using visual dynamics,” *IEEE T INF FOREN SEC*, vol. 10, no. 4, pp. 762-777, 2015. Article (CrossRef Link).

[26] A. d S. Pinto, H. Pedrini, W. R. Schwartz and A. Rocha, “Video-based face spoofing detection through visual rhythm analysis,” in *Proc. of IEEE 25th Conf. on Graphics, Patterns and Images*, pp. 221-228, 2012. Article (CrossRef Link).

[27] A. Anjos et al., “Face anti-spoofing: visual approach,” *Handbook of Biometric Anti-Spoofing*, 1st ed., Springer, London, 2014, Ch. 4, pp. 65-83. Article (CrossRef Link)

[28] G. Pan, L. Sun, Z. Wu and S. Lao, “Eyeblink-based anti-spoofing in face recognition from a generic webcamera,” in *Proc. of IEEE 11th Int. Conf. Computer Vision (ICCV)*, pp. 1-8, 2007. Article (CrossRef Link).

[29] J. Komulainen, A. Hadid and M. Pietikäinen, “Context-based face anti-spoofing,” in *Proc. of IEEE 6th Int. Conf. on Biometrics: Theory, Applications and Systems (BTAS)*, pp. 1-8, 2013. Article (CrossRef Link).

[30] M. I. Faraj and J. Bigun, “Audio-visual authentication using lip-motion from orientation maps,” *Pattern Recognition Letters*, vol. 28, pp. 1368-1382, 2007. Article (CrossRef Link)

[31] Z. Zhang, D. Yi, Z. Lei and S. Z. Li, “Face liveness detection by learing multiplectral reflectance distributions,” in *Proc. of IEEE Int. Conf. on Automatic Face and Gesture Recognition (AFGR)*, pp. 436-441, 2011. Article (CrossRef Link).

[32] I. Pavlidis and P. Symosek, “The imaging issue in an automatic face/disguise detection system,” in *Proc. of IEEE Int. Workshop on Computer Vision Beyond the Visible Spectrum: Methods and Applications*, pp. 15-24, 2000. Article (CrossRef Link).

[33] K. Patel, H. Han, A. K. Jain and G. Ott, “Live face video vs. spoof face video: Use of moiré patterns to detect replay video attacks,” in *Proc. of IEEE Int. Conf. on Biometrics (ICB)*, pp. 98-105, 2015. Article (CrossRef Link).

[34] P. Vuilleumier, J. L. Armony, J. Driver and R. J. Dolan, “Distinct spatial frequency sensitivities for processing faces and emotional expressions,” *Nature Neuroscience*, vol. 6, no. 6, pp. 624-631, 2003. Article (CrossRef Link)

[35] R. Basri and D. W. Jacobs, “Lambertian reflectance and linear subspaces,” *IEEE Trans. on Pattern Analysis and Machine Intelligence*, vol. 25, no. 2, pp. 218-233, 2003. Article (CrossRef Link)

[36] H. J. Nussbaumer, “The fast Fourier transform,” *Fast Fourier Transform and Convolution Algorithms*, 2nd ed., Springer-Verlag, Berlin-Heidelberg, Ch. 4, pp. 80-111, 1982. Article (CrossRef Link)

[37] D. G. Lowe, “Distinctive image features from scale-invariant key points,” *Int. Journal of Computer Vision*, vol. 60, no. 2, pp. 91-110, 2004. Article (CrossRef Link).

[38] Z. Akhtar and G. L. Foresti, “Face spoof attack recognition using discriminative image patches,” *Journal of Electrical and Computer Engineering*, vol. 2016, pp. 1-14, 2016. Article (CrossRef Link).

[39] J. Yang, Z. Lei, S. Liao and S. Z. Li, “Face liveness detection with component dependent descriptor,” in *Proc. of IEEE Int. Conf. on Biometrics (ICB)*, pp. 1-6, 2013. Article (CrossRef Link).

[40] C. C. Chang and C. J. Lin, “LIBSVM: a library for support vector machines,” *ACM Trans. on Intelligent Systems and Technology*, vol. 2, no. 27, pp. 1-27, 2011. Article (CrossRef Link)

[41] S. Bharadwaj, T. I. Dhamecha, M. Vatsa and R. Singh, “Computationally efficient face spoofing detection with motion magnification,” in *Computer Vision and Pattern Recognition Workshops*, pp. 105-110, 2013. Article (CrossRef Link).

[42] D. Menotti et al., “Deep Representations for iris, face, and fingerprint spoofing detection,” in *IEEE T INF FOREN SEC*, vol. 10, no. 4, pp. 864-879, April 2015. Article (CrossRef Link)

[43] N. Erdogmus and S. Marcel, "Spoofing in 2D face recognition with 3D masks and anti-spoofing with kinect," in *Proc. of IEEE 6th Int. Conf. on Biometrics:Theory, Application and Systems (BTAS)*, pp. 1-6, 2013. Article (CrossRef Link)

[44] A. Pinto, W. R. Schwartz, H. Pedrini, A. d. R. Rocha, "Using visual rhythms for detecting video-based facial spoof attacks," *IEEE Trans. on Information Forensics and Security*, vol.10, no.5, pp.1025-1038, May 2015. Article (CrossRef Link)

[45] A. Pinto, H. Pedrini, W. R. Schwartz and A. Rocha, "Face spoofing detection through visual codebooks of spectral temporal cubes," *IEEE Trans on Image Processing*, vol. 24, no. 12, pp. 4726-4740, Dec. 2015. Article (CrossRef Link)

[46] I. Manjani et al., "Detecting silicone mask-based presentation attack via deep dictionary learning," *IEEE T INF FOREN SEC*, vol. 12, no. 7, pp. 1713-1723, July 2017. Article (CrossRef Link)

[47] D. Gragnaniello, G. Poggi, C. Sansone and L. Verdoliva, "An investigation of local descriptors for biometric spoofing detection," *IEEE T INF FOREN SEC*, vol. 10, no. 4, pp. 849-863, April 2015. Article (CrossRef Link)

[48] F. Liu et al., "Local structure based multi-phase collaborative representation for face recognition with single sample per person," *Information Sciences*, vol. 346-347, pp. 198-215, June 2016. Article (CrossRef Link)

[49] Q. Ye et al., "L1-norm distance minimization-based fast robust twin support vector k-plane clustering," *IEEE Trans on Neural Netw. and Learning Syst.*, vol. PP, no. 99, pp. 1-10, 2017. Article (CrossRef Link)

[50] H. Yan et al., "L1-norm GEPSVM classifier based on an effective iterative algorithm for classification," *Neural Processing Letters*, vol. 48, no. 1, pp. 273-298, 2017. Article (CrossRef Link)

[51] Q. Ye et al., "Lp- and Ls-norm distance based robust linear discriminant analysis," *Neural Networks*, vol. 105, pp. 393-404, 2018. Article (CrossRef Link)

[52] F. Liu, J. Tang, and Z. Tang, "An eye states detection method by using WLBP," in *Proc. of 7th IEEE Int. Conf. on Semantic Computing*, CA, pp. 198-201, 2013. Article (CrossRef Link)

[53] F. Liu, Z.Tang, and J. Tang, "WLBP: weber local binary pattern for local image description," *Neurocomputing*, vol. 120, pp. 325-335, November 2013. Article (CrossRef Link)

[54] X. Shu et al., "Image classification with tailored fine-grained dictionaries," *IEEE Trans. on Circuits and Systems for Video Technology*, vol. 28, no. 2, pp. 454-467, Feb. 2018. Article (CrossRef Link)

[55] Z. Li and J. Tang, "Unsupervised feature selection via nonnegative spectral analysis and redundancy control," *IEEE Trans Image Process.*, vol. 24, no. 12, pp. 5343-5355, Dec. 2015. Article (CrossRef Link)

[56] Z. Li, J. Tang and X. He, "Robust structured nonnegative matrix factorization for image representation," *IEEE Trans on Neural Netw. and Learning Syst.*, vol. 29, no. 5, pp. 1947-1960, May 2018. Article (CrossRef Link)

Zahid Ali received his B.E in Electronic Engineering from NED University in 2009 and M.E in Computer Science and Engineering from Chosun University in 2013. He was awarded Global IT Scholarship by Republic of South Korea in 2013. He is currently a Ph.D. candidate in the department of Computer Science and Engineering at Sogang University, where he is also working as research assistant in Computer Vision and Image Processing lab. His main research interests are at the intersection of computer vision and machine learning, developing techniques for unconstrained face recognition and objects detection using Convolutional Neural Networks (CNN's).

Unsang Park received his B.S. and M.S. degrees from the Department of Materials Engineering, Hanyang University, Seoul, Korea, in 1998 and 2000, respectively. He received his M.S. and Ph.D. degrees from the Department of Computer Science and Engineering, Michigan State University, MI, USA in 2004 and 2009, respectively. He has been an assistant professor in the Department of Computer Science and Engineering at Sogang University since 2012. His research interests include pattern recognition, image processing, computer vision,