A Cross-Dataset Evaluation of Anti-Face-Spoofing Methods Using Random Forests and Convolutional Neural Networks

Chin-Shyurng Fahn
Department of Computer Science and Information Engineering
National Taiwan University of Science and Technology
No. 43, Keelung Rd., Sec. 4,
Da'an Dist., Taipei 10607, Taiwan
csfahn@mail.ntust.edu.tw

Chu-Ping Lee
Department of Computer Science and Information Engineering
National Taiwan University of Science and Technology
No. 43, Keelung Rd., Sec. 4,
Da'an Dist., Taipei 10607, Taiwan
D10215011@mail.ntust.edu.tw

Meng-Luen Wu
Department of Computer Science and Information Engineering
National Taiwan University of Science and Technology
No. 43, Keelung Rd., Sec. 4,
Da'an Dist., Taipei 10607, Taiwan
D10015015@mail.ntust.edu.tw

ABSTRACT

Face recognition for authentication, namely unlocking by faces, is widely used in various access control applications, especially in mobile devices, and becomes one of major biometric authentication technology. Some existing authentication methods require additional depth sensors; however, they are still cheated by 2D or 3D printed faces sometimes. Although many researches aim at detecting fake faces, most of them only work well on specific situations, and they are unusable to master unseen spoofed scenarios.

Accordingly, in this paper, we propose face liveliness detection methods using a conventional camera, which is capable of effectively performing both intra- and cross-dataset detection on sets of real faces mixed with spoofed ones. We adopt local binary patterns (LBP) and 2D image distortion analysis (IDA) to extract texture information of face images, which are used for developing our face liveness detection system against spoofing attack to distinguish fake faces from real ones by a deep neural network (DNN). In addition to verifying whether the deep learning method induces over-fitting of spoofed faces using specific datasets, we also employ a random forest classifier to compare the face liveliness detection results. In intra-dataset evaluation, 10-fold cross-validation is adopted, and the accuracy of spoofed face detection is more than 97% using a convolutional neural network architecture. In cross-dataset evaluation, under the condition of the Idiap Replay-Attack Database acting as the training dataset as well as the NUAA Photograph Imposter Database serving as the testing dataset, the accuracy achieves 81.85% when using the scheme of combining LBP, IDA, and DNN techniques. Such performance is better than state-of-the-art methods.

CCS Concepts

• Security and privacy → Intrusion/anomaly detection and malware mitigation → Intrusion detection systems.

AICCC 2019, December 21–23, 2019, Kobe, Japan

ACM ISBN 978-1-4503-7263-3/19/12…$15.00
DOI: https://doi.org/10.1145/3375959.3375985

Keywords

face authentication; spoof attack detection; random forest; deep neural network; convolutional neural network; cross-dataset evaluation; local binary pattern; image distortion analysis.

1. INTRODUCTION

In recent years, digital privacy has become a major issue, and biometric authentication methods are developed to prevent unauthorized use of personal electronic devices, and perform access control on secret and confidential data. There are many biometric authentication methods, including retinal recognition, fingerprint identification, voiceprint identification, handwriting recognition, and face recognition. Particularly, face recognition is an intuitive method that the user can complete an authentication process simply by looking at a mobile phone, and the frontal camera captures the necessary information simultaneously and seamlessly in the background. Therefore, face recognition for authentication, also called "unlocking by faces," has become one of the most popular methods in the field of biometric authentication

However, there are some issues for authentication by faces. The first one is the spoofing attack. For example, reproducing an authorized user's face photo can cheat an authentication system. Replaying victim's face video, or making a 3D face sculpture of the victim, is still efficacious for cracking an authentication system. The second issue is system construction costs. For higher recognition accuracy and spoofing attack prevention, some face authentication systems adopt additional sensors rather than conventional cameras, such as an infrared camera, a flood illuminator, and a dot projector. The requirement of these equipment also limits the portability of the face authentication systems that only need a facial camera.

To brief, an effective anti-face spoofing attack method can avoid malicious operations and lower system construction costs. Currently, face liveliness detection is one of the prominent methods for anti-face spoofing attack. In the state-of-the-art camera based liveliness detection methods, they operate well on specific environments, but are not available on some conditions. For example, with these methods, the accuracy is high on a face spoofing dataset, but unacceptably low on other datasets. In consequence, to improve the effectiveness of existing anti-face spoofing attack methods, we propose a face liveliness detection method based on digital images captured by frontal face cameras, and increase its generalization ability for multiple face spoofing datasets, including our own built dataset.

2. RELATED WORK

In this section, we introduce the related work about face liveliness detection, inclusive of texture based, motion based, image quality analysis based, and depth camera based approaches. We also elaborate the pros and cons of these various types of spoofed face detection methods.

2.1 Texture Based Methods

The texture based method analyzes the texture details of face images, which are caused by skin roughness, local light, and so on. Yang et al. extracted LBP and HOG texture features to distinguish fake faces from real ones [1]. In their experimental results, they found that texture based methods are effective to distinguish between fake and real face images in both the CASIA and Idiap databases. By adding texture clues, The Half Total Error Rate (HTER) on the Idiap database decreased by 13.87% in [2], 7.60% in [3], and 6.62% in [4].

However, the generalization ability of texture based methods is poor. A study reported in [5] shows that for two texture based methods in [2] and [3], the HTER increased significantly in the scenario of cross-dataset evaluation; that is, testing and training datasets come from different face spoof databases. Because of the inherent data characteristics of texture-based methods, they are easily over-fitted on a particular illumination and image resolution condition. As a result, the texture based methods do not manipulate well for various illumination and image resolutions.

2.2 Motion Based Methods

The motion based methods work on detecting the movement of organs and facial muscles, such as blinking, mouth movement [6], and head twist [7]. Motion features are originated from motion information in different time series of a film. Compared to texture based methods, this kind of methods has better generalization ability. Nevertheless, there are also some limitations of motion based methods. For example, facial movements are confined to human physiology, ranging from 0.2 Hz to 0.5 Hz [4]. Therefore, it takes a relatively long time to acquire life features for face spoofing detection, usually more than 3 seconds. In addition, motion based methods are easily influenced by external factors, leading to discriminatory problems, such as dynamic background.

2.3 Methods Based on Image Quality Analysis

In 2014, Galbally et al. proposed a method based on image quality assessment to enhance the security of biometrics recognition [8]. The image quality assessment has 21 full-reference measures and 4 non reference measures, including Error Sensitive Measures (ESM), Structural Similarity Measures (SSM), Information Theoretic Measures (ITM), Distortion Specific Measures (DSM), Training Based Measures (TBS), and Natural Scene Statistic Measures (NSSM). The goal of this method is to develop a living body detection manner across different biometric modes. The advantages are that only one image is required to distinguish the authenticity and the computational cost is low. The disadvantage is the lack of generalization ability for varied datasets.

2.4 Methods Based on Depth Cameras

Some face spoof detection methods perform on a depth camera and use the depth image to classify whether an image contains real face or not. There are many types of depth cameras, and the most common one is equipped with Near-infrared (NIR) lenses. The spectral band of NIR is different from that of visible light, so the absorption and reflection intensities of real faces and spoofing faces in the NIR band are dissimilar. Hence, NIR images can be used in face liveness detection. It is possible to extract texture features from NIR images [9] that have a greater discrimination against screen attacks and a smaller discrimination against high resolution color printed paper. The main disadvantage of this kind of methods is that additional depth cameras are required.

3. FACE LIVENESS DETECTION

In this paper, face liveness detection is the anti-face spoofing solution. There are two phases of analysis in the detection. The first phase is image texture analysis in which Local Binary Pattern (LBP) is selected as the cue, because the textures of the face image captured directly from a real face has subtle differences from those reproduced from a printer. The second phase is image distortion analysis, which is based on an assumption that any reproduced image has lower quality in colors and sharpness. In this phase, a series of Image Distortion Analysis (IDA) is applied.

3.1 Local Binary Pattern

LBP is a texture descriptor, which calculates the local representation of textures [10]. This local representation is realized by comparing each pixel with its surrounding neighboring pixels. In Figure 1, the LBP is generated from setting the intensity of the central pixel as a threshold used for altering the intensities of its eight neighboring pixels into binary values. The intensity of a neighboring pixel is set to 1, if it is greater than or equal to the threshold; otherwise, set to 0. For eight neighboring pixels, we have a total of 256 possible combinations of LBP codes.

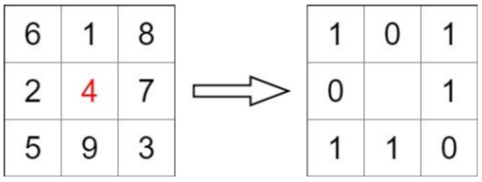


Figure 1. An example of generating the LBP code.

The following depicts how to obtain an LBP value. In a 3 x 3 window, there are 8 neighbors around the center. We can start from any neighboring pixel and iterate all the neighbors clockwise or counter-clockwise consistently. For instance, we start at the left upper pixel and iterate all neighboring pixels in a clockwise order, which is stored in an 8-bit binary code. Figure 2 shows the binary code converted to a decimal value, say 182. This is called the LBP value.

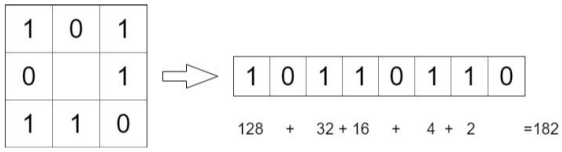
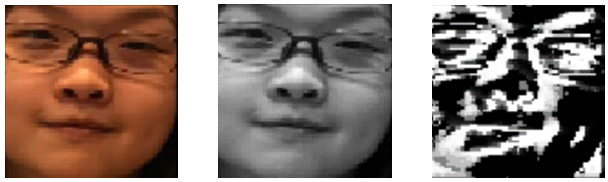


Figure 2. An example of calculating the LBP value.

Through repeating the aforementioned process for each pixel in an image, the resulting LBP image is obtained as an example shown in Figure 3. From this figure, we can see that both the variation of texture orientations and smoothness of a real face image are obviously reserved.



(a) (b) (c)

Figure 3. Illustration of an input image resulting in an LBP image: (a) the original image; (b) the grayscale image; (c) the resulting LBP image.

3.2 Image Distortion Analysis

In this section, four different features of IDA, which comprises specular reflection, sharpness, chromatic moment, and color diversity, are described as follows.

3.2.1 Specular Reflection Features

When the incident light is reflected from a multitude of angles, diffuse reflection occurs. In this case, the incident energy is distributed in all directions of reflection. It usually happens at the contact point of a particle surface. This is a reflection that allows us to see objects and their shapes. When the incident light is reflected in only one direction, namely the specularity, also known as specular reflection occurs. In this case, the incident and reflected light have the same energy, and there is no loss in principle. This energy can emit light, especially when the light source is near the surface. This reflection occurs when the surface is smooth. Reflections usually have both specular and diffuse components as Figure 4 shows.

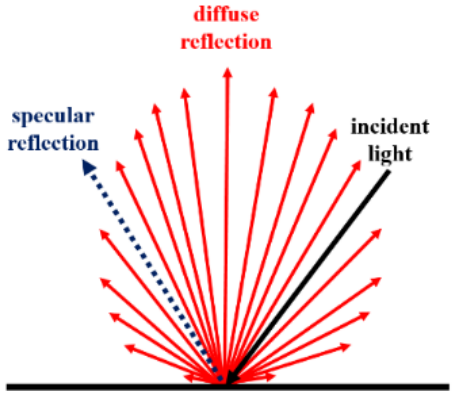


Figure 4. Diffuse and specular reflections.

A spoofed face image usually appears like a mirror reflection. The analysis of mirror reflection components has been widely used for standardization of facial lighting [11] and removal of mirror reflection [12]. We adopt a 2D histogram to acquire the specular reflection component [13] from an input image. The 2D histogram is in form of the following equations:

$$i = \frac{1}{3}(r + g + b) \tag{1}$$

$$s = \begin{cases} \frac{1}{2}(2r - g - b) = \frac{3}{2}(r - i), if\,(b + r) \geq 2g \\ \frac{1}{2}(r + g - 2b) = \frac{3}{2}(i - b), if\,(b + r) < 2g \end{cases} \tag{2}$$

where i is the intensity, s is the saturation, and r, g, and b represent the red, green, and blue components of an image, respectively. The specular reflection can be identified from a 2D histogram of i and s. A pixel p is regarded as a part of the mirror area if it satisfies the following conditions:

$$p = \begin{cases} 0, & if\ i_p \geq \frac{1}{2} i_{max} \text{ and } s_p \leq \frac{1}{3} s_{max} \\ 1, & otherwise \end{cases} \tag{3}$$

where i_{max} and s_{max} stand for the maximum intensity of i and the maximum saturation of s for all pixels in a single image, respectively. After all pixels have been examined, a binary mask of specular reflection is obtained. And we compute the ratio of the mask to the input image as a feature for face liveness detection. An example of detecting specular reflection is illustrated in Figure 5.

(a) (b) (c)

Figure 5. Illustration of detecting specular reflection: (a) the original image; (b) the binary mask of specular reflection; (c) the resulting image from reflection removal.

After reflection removal, we convert the resulting image into a gray-scale image, and find its average and standard deviation of intensity as the other two features for detecting spoofed face images.

3.2.2 Sharpness Features

Because the shooting capability of some face cameras is low for the short distance from targets, the captured images are often out of focus and become blur. In addition, the resolution of a printer or monitor display is sometimes lower than that of the original image. As a result, compared to real face images, spoofed face ones tend to be poorer quality in terms of sharpness. To evaluate the sharpness of an image, we apply the Laplacian operator [14] to measuring the second derivative of an image. The Laplacian operator highlights the areas of rapid intensity changes, which is often used for edge detection, like Sobel and Scharr filters.

If an image contains high variance of responses made by Laplacian filtering, including both edge-like and non-edge like ones, the image is often in-focus; otherwise, it is out of focus, and is supposed to be a spoofed image. The second derivative operator is employed to keep high spatial frequency pixels and remove low frequency ones. The second derivative of an image can be obtained through convoluting the image with the Laplacian operator. Such an operator represented by the Laplacian mask is stated below:

$$M = \begin{bmatrix} 0 & -1 & 0 \\ -1 & 4 & -1 \\ 0 & -1 & 0 \end{bmatrix} \tag{4}$$

Figure 6 graphically shows the Laplacian convolution result from an input image where most of facial features are detected.

(a) (b) (c)

Figure 6. Illustration of Laplacian convolution: (a) the original image; (b) the grayscale image; (c) the resulting convoluted image.

After obtaining the edge image, we calculate the variance of the absolute Laplacian convoluted values by the following equation:

$$VAR(L) = \sum_{i=1}^{m}\sum_{j=1}^{n}(|L(i,j)| - \bar{L})^2 \quad (5)$$

where $L(i,j)$ is the convolution of the input image I with the mask M. And $\bar{L}$ is the mean of absolute values of $L(i,j)$, which yields:

$$\bar{L} = \frac{1}{mn}\sum_{i=1}^{m}\sum_{j=1}^{n}|L(i,j)| \quad (6)$$

The above mentioned variance of the absolute Laplacian convoluted values also plays the role of a feature for face liveness detection.

3.2.3 Chromatic Moment Features

Color reproducibility on a screen is different from that in paper, and recaptured face images and real face images are often dissimilar in color distribution. Accordingly, we use chromatic moments to analyze the color distribution of an image. Each chromatic moment is an effective color feature proposed by Stricker and Orengo [15], which is based on the mathematical statistics that any color distribution of an image can be represented by its moments. In general, low-order moments, such as the first-order, second-order, and third-order moments, are sufficient to express the color distribution of an image. In this paper, we calculate the first-order, second-order, and third-order moments of all pixels' colors in each color channel. Hence, the chromatic moments of an image have 9 components (3 color channels and 3 low-order moments per channel), which are formulated as follows.

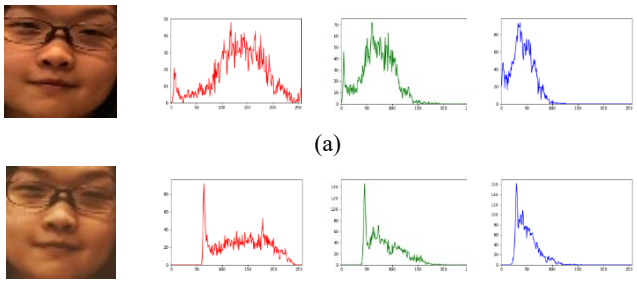
$$\begin{cases} M_{k1} = E_k = \dfrac{1}{mn}\displaystyle\sum_{i=1}^{m}\sum_{j=1}^{n} p_{ij}^k \\ M_{k2} = \sigma_k^2 = \dfrac{1}{mn}\displaystyle\sum_{i=1}^{m}\sum_{j=1}^{n}\left(p_{ij}^k - E_k\right)^2 \\ M_{k3} = S_k = \dfrac{1}{mn}\displaystyle\sum_{i=1}^{m}\sum_{j=1}^{n}\left(p_{ij}^k - E_k\right)^3 \end{cases} \quad k = 1,2,3 \quad (7)$$

where k is the channel index. The above three formulas are equivalent to the lower-order moments in each color channel. Moreover, we calculate the standard deviation and the third root of the skewness of all pixels' colors in each color channel.

For a face image, we first transform the RGB color space into the HSV one. The advantage of this HSV color space is that Hue is not sensitive to the change of light, and is not influenced by brightness contrast and white light reflection. Then the average intensity, deviation, and skewness of each HSV channel act as chromatic moment features.

3.2.4 Color Diversity Features

Recaptured images often lose color details. As can be seen from Figure 7, real face images have richer colors than spoofed ones do. To obtain the degree of color diversity, we select 10 pixel values with the highest repetition in the R, G, and B color channels, respectively, totally 30 color diversity features.



(a)

(b)

Figure 7. Color diversity in RGB histograms: (a) a real face image; (b) a spoofed face image captured from a printed photo.

4. SPOOFED FACE IMAGE SCREENING

In this section, three methods for the detection of spoofed faces are described. The first and the second methods are based on the LBP, IDA, deep neural network (DNN), and random forest (RF). To evaluate the performance of different schemes, the third method is to develop a convolutional neural network (CNN) for screening spoofed face images.

4.1 Random Forest

The RF is a supervised classification and regression learning method [16], which builds multiple decision trees and combines their outputs to improve the generalization ability of the existing spoofed face detection schemes. There are hundreds of decision trees in an RF, and the final results are determined by a majority vote of the outputs of the trees [17]. Suppose an RF consists of n classifiers in form of decision trees denoted as:

$$h(X,\theta_k), \quad k = 1,2,\dots,n \quad (8)$$

where θ_k is a random variable with a given independent and identical distribution, and X is an input vector. After sending an input X to the RF, each decision tree outputs a class regarded as a vote, and the majority vote is the classification result of X. The random variable θ_k is used to control the growth of the k^{th} decision tree that is generated from a training dataset. Then a classifier $h(X,\theta_k)$ is created by the aid of θ_k, where X is an input vector. After that, we simplify $h(X,\theta_k)$ to $h_k(X)$. Given a series of classifiers $h_1(X), h_2(X), \dots, h_n(X)$, the margin function is formulated as follows:

$$mg(X,Y) = av_k I(h_k(X) = Y) - \max_{j \neq Y} av_k I(h_k(X) = j) \quad (9)$$

where X is a sample vector, Y is a classification vector, $I(\cdot)$ is an indicator function, and $av(\cdot)$ is an average function. The margin function represents the degree of the votes of X greater than that of any other failed classifier Y. The larger the margin is, the higher the confidence of the classification obtains.

The key of establishing an RF is how to select the optimal m [18] by Out-of-Bag (OOB) estimation processing [19],[20]. With the RF method, the optimal segmentation of a decision tree T is determined via the Gini index [21], which is described below:

where N is the number of classes, and P_j is the probability of training samples that belong to class j. If the decision tree T is partitioned into m subtrees, then the split version of the Gini index can be written as a finite series of m terms:

$$gini(T) = 1 - \sum_{j=1}^{N} P_j^2 \quad (10)$$

where the numbers of classes categorized from subtrees T_1 to T_m are N_1 to N_m in order. The rule of selecting variable m is accomplished by taking the smallest $gini_{split}$ when generating a classification and regression tree.

$$gini_{split}(T) = \frac{N_1}{N} gini(T_1) + \cdots + \frac{N_m}{N} gini(T_m) \quad (11)$$

4.2 Deep Neural Network

DNN is a type of multilayer perceptron that contains at least more than four hidden layers [22]. With more hidden layers, the network is able to solve the classification problems in higher dimensions. Herein, we use a total of 256 LBP and 43 IDA features (specularity: 3, sharpness: 1, chromatic moment: 9, color diversity: 30) as the input for our DNN which consists of five fully connected network layers as shown in Figure 8. After multi-layer calculation of the DDN, the prediction probabilities of the true face and the false face will be acquired finally.

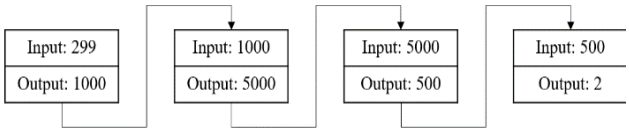


Figure 8. Our DNN architecture consisting of five fully connected network layers.

4.3 Convolutional Neural Network

In addition to the DNN, we also apply a CNN to screening spoofed face images. The main purpose of using the CNN is to extract facial features from input face images automatically. The design philosophy of our CNN architecture is originated from VGG16 [23], as Figure 9 illustrates. This architecture includes four 3x3 convolution layers used for feature extraction and three fully-connected network layers. In addition, each convolution layer is followed by one pooling layer. The four pooling layers divide the CNN architecture into five blocks. The first four blocks comprising four pairs of convolution and pooling layers are used to extract image features, and the last block is composed of three fully connected network layers, where the feature map is transformed into one-dimensional data used for classification.

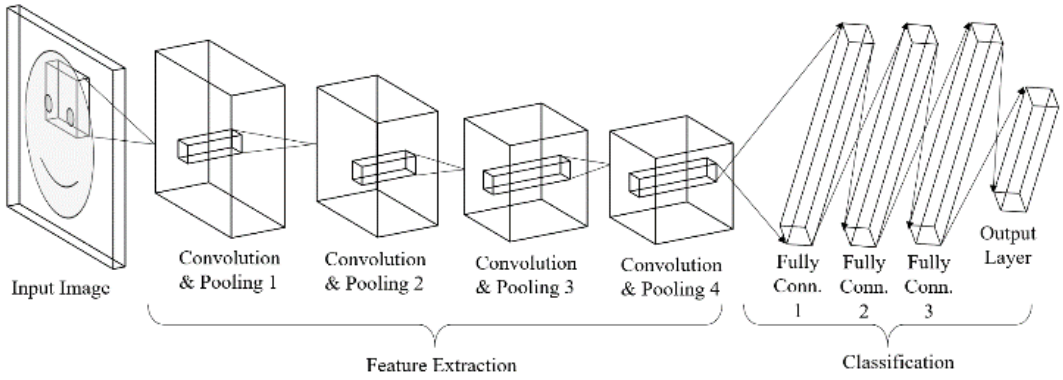


Figure 9. Illustration of our CNN architecture.

5. EXPERIMENTAL RESULTS AND DISCUSSION

In this section, we elaborate and analyze the experimental results from both the intra-dataset and cross-dataset evaluation of anti-face-spoofing methods. First, we briefly depict the experimental setup. Second, three databases used for face spoofing attack, including our own built database, are described. Third, the results of intra-dataset face liveness detection are presented. At last, we show the cross-dataset spoofed face detection results.

5.1 Experimental Setup

Our face liveness detection system runs on a personal computer, and its main specifications are Intel® Core™ i5-6500 @ 3.2 GHz and 16GB RAM. We use the Mirrorless Interchangeable Lens Camera (MILC) OLYMPUS EM-5 Mark II to capture images whose sizes are 4608×3456. Table 1 lists the hardware and software devices used in the experiments.

Table 1 The experimental setup in our system

Tool	Version or Specification
Computer Hardware	CPU: Intel® Core™ i5-6500 @ 3.2 GHz GPU: NVIDIA GeForce GTX1060 6GB RAM: 16GB DDRIII
Operating System	Microsoft Windows 10 64-bit
Developing Tools	Python 3.6 OpenCV 3.4.1

5.2 Face Spoof Databases

In order to evaluate the effectiveness of the face liveness detection system, we use some proprietary spoofing databases, including the NUAA Photograph Imposter Database [24] and the Idiap Replay-Attack Database [2].

5.2.1 NUAA Photograph Imposter Database

The NUAA Photograph Imposter Database was released in 2010 and is one of the earliest public domain face spoofing databases, which consists of 12,614 face images selected from 143 videos, and 15 subjects are invited to attend in this work. The images are

captured by a conventional IP camera. The acquisition process includes various changes of facial poses. Besides this, such images cover the face samples of different lighting environments, different genders, different ages, and taken from different time. The size of each image is 640×480. In addition, the NUAA database contains only printed photo attacks.

5.2.2 Idiap Replay-Attack Database

The Idiap Replay-Attack Database is published by the Swiss Idiap research institute, which comprises 1,300 color video recordings, including real-life and spoofing attack access from 50 different subjects. These videos are captured with a webcam under the same scenario that the user is performing a verification attempt by either a real person or a generated face spoofing attack. Each real person has a spoofed video version, and the length of a video is at least 9 seconds long. In the spoofing attack, the testing faces are either from printed photos or displayed photos and videos. The videos with a resolution of 320×240 pixels were recorded on a Macbook laptop using the QuickTime framework in Motion JPEG format. The frame rate is about 25 Hz. In our experiment, the videos are converted into 9,797 images.

5.2.3 Our Own Built Database

To test our anti-face-spoofing methods, we also have built an own database, including 1,856 real face images and 1,603 spoofed face images, totally 3,459 images. Distinguishing from the images of the NUAA and Idiap databases, our real face images are shot by a mobile phone. And the spoofed face images are recaptured by an MILC from printed photos and Full-HD display, which also simulates the exposure environment with different lighting conditions.

5.3 Results of Intra-dataset Spoof Detection

In this section, we perform an intra-dataset evaluation using the three databases, including NUAA Photograph Imposter Database, Idiap Replay-Attack Database, and our own built database. Each database is evaluated separately, which is partitioned into a training dataset and a testing dataset. To avoid bias on a particular partition, we adopt 10-fold cross-validation to evaluate the accuracy of spoofed face detection.

Table 2 shows the experimental results of face liveness detection using our proposed schemes, including Raw images+CNN, LBP+IDA+DNN and LBP+IDA+RF. They are compared to the other two schemes, DoG+SNLR [24] and LBP+SVM [2]. The experimental result reveals that our anti-face-spoofing methods have the higher detection accuracy. For both the NUAA and our own built databases, the Raw images+CNN scheme performs the best, while for the Idiap database, the LBP+IDA+DNN scheme is the prime choice.

Table 2. The accuracy of spoofed face detection in an intra-dataset for different schemes

Database / Scheme	NUAA	Idiap	Ours
DoG+SNLR [24]	86.70%	N/A	N/A
LBP+SVM [2]	80.97%	84.84%	N/A
Raw images+CNN (Our proposed)	99.99%	97.87%	98.72%
LBP+IDA+DNN (Our proposed)	95.44%	99.55%	94.10%
LBP+IDA+RF (Our proposed)	99.04%	98.38%	97.75%

5.4 Results of Cross-dataset Spoof Detection

Besides the intra-dataset evaluation, the cross-dataset face spoof detection is also evaluated under the condition that training and testing datasets are given by different databases. The metric formula of detection accuracy is defined below.

$$Accuracy = \frac{TP + TN}{TP + FN + FP + TN} \tag{12}$$

In the above formula, the positive samples (spoofed faces) that are correctly labeled by the classifier are called true positives (TP). The negative samples (real faces) that are correctly labeled by the classifier are called true negatives (TN). The negative samples that are incorrectly labeled as spoofed faces are called false positives (FP). The positive samples that are mislabeled as real faces are called false negatives (FN). Table 3 reveals the experimental results of the spoofed face detection using our proposed schemes including Raw images+CNN, LBP+IDA+RF, and LBP+IDA+DNN. In this experiment, the Idiap database acts as the training dataset and both the NUAA and our own built databases serve as the testing dataset.

Table 3. The accuracy of cross-dataset spoofed face detection using the Idiap database as the training dataset

Testing Dataset / Scheme	NUAA	Ours
Raw images+CNN (Our proposed)	45.74%	57.18%
LBP+IDA+DNN (Our proposed)	81.85%	60.34%
LBP+IDA+RF (Our proposed)	65.47%	62.91%

From this table, it can be observed that the accuracy of spoofed face detection using the Raw images +CNN scheme is much lower than that of the other two because the CNN architecture cannot extract the distinct features of real faces and spoofed faces. As expected, the performance of cross-dataset evaluation is inferior to that of intra-dataset evaluation. The failure examples for spoofed face detection are shown in Figure 10 where the images are provided by the NUAA database.

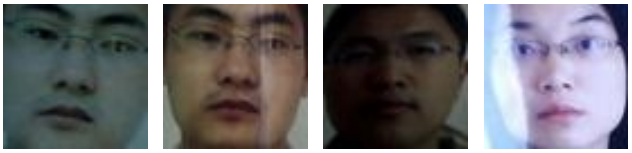


Figure 10. Examples of incorrectly detecting the spoofed face images in the NUAA database.

To sum up, under the diversified shooting environment, the latter two schemes in the table have higher accuracy than the first one does, which means that they possess preferable generalization ability of detecting spoofed faces in case of training and testing across different databases.

6. CONCLUSIONS

In this paper, a deep-learning-based face liveness detection system is presented against spoofing attack using 2D features of LBP and IDA based on the classifiers of RF and DNN, which can effectively improve the performance of information security systems by virtue of face authentication techniques.

First, we propose a feature extraction method used for face liveness detection against spoofing attack. In this method, five kinds of features are extracted, which comprises 256 LBP values combined with four types of 43 IDA features (specularity: 3, sharpness: 1, chromatic moment: 9, and color diversity: 30), resulting in a 299-dimensional feature vector. Accordingly, a DNN architecture is designed as a classifier to screen spoofed face images. Moreover, we adopt a RF as a traditional machine learning model to compare with the DNN architecture on the ability of distinguishing between real faces and spoofed ones. We also create a CNN architecture to classify raw face images and evaluate its performance of face liveness detection.

Finally, our own built database is applied in the experiment, and its constituting real face images are taken by a mobile phone, whereas spoofed face images are recaptured by an MILC from printed photos and Full-HD display. In the intra-dataset spoofed face detection, not all the performance resulting from using the Raw images+CNN scheme is the best. By contraries, the accuracy of spoofed face detection is up to 99.55% using the LBP+IDA+DNN scheme with 10-fold cross-validation on the Idiap database, which performs better than that of the schemes using DoG+SNLR, LBP+SVM, and Raw images+CNN. As for the cross-dataset spoofed face detection, the accuracy achieves 95.13% using the LBP+IDA+RF scheme, which is superior to that using the LBP+IDA+DNN one. It is greatly disappointed that the Raw images+CNN scheme obtains the worst performance.

7. ACKNOWLEDGEMENTS

The authors thank the Ministry of Science and Technology of Taiwan (R. O. C.) for supporting this work in part under Grant MOST 107-2221-E-011-113-MY2. And also thank Mr. Tzu-Yuan Wu for enthusiastically participating the development of the anti-face-spoofing methods.

8. REFERENCES

[1] Yang, J. and Li, S. Z. 2013. Face liveness detection with component dependent descriptor. In *Proceedings of the International Joint Conference on Biometrics* (Madrid, Spain, June 04-07, 2013). IJCB'13. IEEE, New York, NY, 1-6. DOI= http://dx.doi.org/10.1109/ICB.2013.6612955.

[2] Chingovska, I., Anjos, A., and Marcel, S. 2012. On the effectiveness of local binary patterns in face anti-spoofing. In *Proceedings of the International Conference of Biometrics Special Interest Group* (Darmstadt, Germany, September 06-07, 2012). BIOSIG'12. IEEE, New York, NY, 1-7.

[3] De Freitas Pereira, T., Anjos, A., De Martino, J. M., and Marcel, S. 2012. LBP-TOP based countermeasure against face spoofing attacks. In *Proceedings of the 11th Asian Conference on Computer Vision* (Daejeon, Korea, November 05-09, 2012). ACCV'12. Springer, Heidelberg, Germany, 121-132. DOI= http://dx.doi.org/10.1007/978-3-642-37410-4_11.

[4] Bharadwaj, S., Dhamecha, T. I., Vatsa, M., and Singh, R. 2013. Computationally efficient face spoofing detection with motion magnification. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops* (Portland, Oregon, June 23-24, 2013). CVPR'13. IEEE, New York, NY, 105-110. DOI= http://dx.doi.org/10.1109/CVPRW.2013.23.

[5] De Freitas Pereira, T., Anjos, A., De Martino, J., and Marcel, S. 2013. Can face anti-spoofing countermeasures work in a real world scenario? In *Proceedings of the International Conference on Biometrics* (Madrid, Spain, June 04-07, 2013). ICB'13. IEEE, New York, NY, 1-8. DOI= http://dx.doi.org/10.1109/ICB.2013.6612981.

[6] Kollreider, K., Fronthaler, H., Faraj, M. I., and Bigun, J. 2007. Real-time face detection and motion analysis with application in liveness assessment. *IEEE Trans. on Info. Foren. and Secur.*, 2, 3 (Aug. 2007), 548-558. DOI= http://dx.doi.org/10.1109/TIFS.2007.902037.

[7] Bao, W., Li, H., Li, N., and Jiang, W. 2009. A liveness detection method for face recognition based on optical flow field. In *Proceedings of the International Conference on Image Analysis and Signal Processing* (Taizhou, China, April 11-12, 2019). IASP'09. IEEE, New York, NY, 233-236. DOI= http://dx.doi.org/10.1109/IASP.2009.5054589.

[8] Galbally, J., Marcel, S., and Fierrez, J. 2013. Image quality assessment for fake biometric detection: application to iris, fingerprint, and face recognition. *IEEE Trans. on Image Proc.*, 23, 2 (Nov. 2013), 710-724. DOI= http://dx.doi.org/10.1109/TIP.2013.2292332.

[9] Sun, X. 2016. Context based face spoofing detection using active near-infrared images. In *Proceedings of the International Conference on Pattern Recognition* (Cancun, Mexico, December 04-08, 2016). ICPR'16. IEEE, New York, NY, 4262-4267. DOI= http://dx.doi.org/10.1109/ICPR.2016.7900303.

[10] Ojala, T. and Pietikainen, M. 2002. Multiresolution gray-scale and rotation invariant texture classification with local binary patterns. *IEEE Trans. on Pat. Anal. and Mach. Intell.* 24, 2 (Aug. 2002), 971-987. DOI= http://dx.doi.org/10.1109/TPAMI.2002.1017623.

[11] Christlein, V. et al. 2013. The impact of specular highlights on 3D-2D face recognition. In *Proceedings of the International Conference on Biometric and Surveillance Technology for Human and Activity Identification* (Baltimore, Maryland, June 03-06, 2013). SPIE, Bellingham, WA, 8712-8719.

[12] Yang, Q., Wang, S., and Ahuja, N. 2010. Real-time specular highlight removal using bilateral filtering. In *Proceedings of the European Conference on Computer Vision* (Heraklion, Crete, September 06-09, 2010). ECCV'10. Springer, Heidelberg, Germany, 87-100.

[13] Tchoulack, S., Pierre Langlois, J. M., and Cheriet, F. 2008. A video stream processor for real-time detection and correction of specular reflections in endoscopic images. In *Proceedings of the International IEEE Northeast Workshop on Circuits and Systems* (Montreal, Canada, June 22-25, 2008). NEWCAS'08. IEEE, New York, NY, 49-52. DOI= http://dx.doi.org/10.1109/NEWCAS.2008.4606318.

[14] Pech-Pacheco, J. L. and Cristóbal, G. 2000. Diatom autofocusing in brightfield microscopy: a comparative study. In *Proceedings of the International Conference on Pattern Recognition* (Barcelona, Spain, September 03-07, 2000). ICPR'00. IEEE, New York, NY, 314-317. DOI= http://dx.doi.org/10.1109/ICPR.2000.903548.

[15] Stricker, M. and Orengo, M. 1995. Similarity of color images. *Stor. and Retr. for Imag. and Vid. Database III*, 2420, 1 (Mar. 1995), 381-392. DOI= http://dx.doi.org/10.1117/12.205308.

[16] Breiman, L. 2001. Random forests. *Mach. Learn.*, 45, 1 (Oct. 2001), 5-32. DOI= http://dx.doi.org/10.1023/A:1010933404324.

[17] Oshiro, T. M., Perez, P. S., and Baranauskas, J. A. 2012. How many trees in a random forest? In *Proceedings of the 8th International Conference on Machine Learning and Data Mining in Pattern Recognition* (Berlin, Germany, July 13-20, 2012). MLDM'12. Springer, Heidelberg, Germany, 154-168.

[18] Genuer, R., Poggi, J. M., and Tuleau-Malot, C. 2010. Variable selection using random forests. *Patt. Recog. Lett.* 31, 14 (Oct. 2010), 2225-2236. DOI= http://dx.doi.org/10.1016/j.patrec.2010.03.014.

[19] Breiman, L. 1996. Bagging predictors. *Mach. Learn.* 24, 2 (Aug. 1996), 123-140. DOI= http://dx.doi.org/10.1023/A:1018054314350.

[20] Breiman, L. 1996. *Out-of-bag estimation*. Technical Report. Department of Statistics, University of California at Berkeley. Retrieved June 20, 2019 from https://www.stat.berkeley.edu/~breiman/OOBestimation.pdf.

[21] Rokach, L., Maimon, O., and Stone C. J. 2005. Top-down induction of decision trees classifiers— a survey. *IEEE Trans. on Sys., Man, and Cyber., Part C*, 35, 4 (Oct. 2005), 476-487. DOI= http://dx.doi.org/10.1109/TSMCC.2004.843247.

[22] Gardner, M. and Dorling, S. 1998. Artificial neural networks the multilayer perceptron: a review of applications in the atmospheric sciences. *Atmos. Environ.*, 32, 14 (Aug. 1998), 2627-2636. DOI= http://dx.doi.org/10.1016/S1352-2310(97)00447-0.

[23] Simonyan, K. and Zisserman, A. 2015. Very deep convolutional networks for large-scale image recognition. In *Proceedings of the International Conference on Learning Representations* (San Diego, California, May 07-09, 2015). ICLR'15. IEEE, New York, NY, 1-14.

[24] Tan, X., Li, Y., Liu, J., and Jiang, L. 2010. Face liveness detection from a single image with sparse low rank bilinear discriminative model. In *Proceedings of the European Conference on Computer Vision* (Heraklion, Crete, September 05-11, 2010). ECCV'10. Springer, Heidelberg, Germany, 504-517. DOI= http://dx.doi.org/10.1007/978-3-642-15567-3_37.