CorneaReflect：Face Liveness Detection for Smartphone Based on Corneal Reflection

WU Bozhi　LI Jingwei　WEN Weiping

(School of Software & Microelectronics, Peking University, Beijing 100871)

Abstract：Nowadays, face recognition technology has been increasingly used on smartphones. But it is vulnerable, and easy to be attacked by presentation attacks, which disrupt the face recognition system by presenting a facial biometric artifact including photo, video replay, and 3D masks. Liveness detection is considered an effective method to detect and mitigate such attacks, and plenty of algorithms based on liveness detection have been proposed. However, most of these algorithms are very complex and not feasible. In order to solve this problem, we propose CorneaReflect, a practical and robust face liveness detection algorithm for smartphone to counter photo-based and video-based attacks, which performs in face liveness detection by judging whether the cornea correctly reflects the contents of the screen of the mobile phone. It does not require any additional hardware or any user interaction. Experimental results show that CorneaReflect can effectively detect photo-based attack and video-based attack, and also has a considerable success rate for the face authentication under different situations.

Key Words：face recognition; presentation attack detection; corneal reflection

The human face, like fingerprints, is a unique biological feature that can be used for identification. In recent years, face recognition technology has become increasingly used in the field of consumer electronics, especially in smartphones. For example, this technology has been used to unlock the smartphone screen and do mobile payment. What is more, face recognition technology has been gradually applied in business, such as enterprise employee management, remote account opening, online medical registration, online payment etc. But face recognition technology is vulnerable, as attackers can deceive the face recognition

作者简介：吴勃志（1990—），男（汉），籍贯（广东省），硕士；李经纬（1995—），男（汉），籍贯（辽宁省），硕士；

通信作者：文伟平，博士，教授，weipingwen@pku.edu.cn

system by demonstrating a false authentication evidence like photos, videos and 3D mask in front of the sensor of face recognition system [1]. To combat such attacks, many scholars have conducted in-depth research and published a large number of research papers. One effective method is facial liveness detection, which is used to verify whether the face presented in front of the face recognition system sensor is a living body, not a photo, video or 3D mask.

At first, people use some simple methods to perform the liveness detection, like challenge response approach [13,14,15,16,17]. It requires users to interact with the system in real time. The blink response is one of the typical approaches of challenge response. However, a well-known spoofing occurred in 2012 intrude the system requiring the user to blink. The adversary spoofed the system merely by an eyes-opening photo and an eye-closing photo [12].

In order to make the attack against face recognition more difficult, people propose the hybrid approaches. These approaches involve face and other biometric traits in the verifying mechanism. One of them purposed to conduct fingerprint authentication at the same time [26]. Another researcher enhanced the system by facial thermo-gram and generic face detection mechanism during the experiment [2, 3]. What is more, the inertial sensors and the head pose changes involved into the face authentication [4]. However, these solutions can't solve the problem radically, which once showed in the research that attackers can invade the system by even a single biometric trait [27].

Some more advanced liveness detection techniques are proposed against the attacks. One of them is texture-based approaches, which are based on analyzing microtextural patterns in the facial image samples [24]. As postulated, the real surface properties differentiate from the counterfeited pigments. Some study contrived micro-texture and a linear SVM to counter counterfeit spoofing [5]. Some studies have been validated by local binary detection, but this type of method requires a very ideal experimental environment and complex algorithm requirements [6]. A multispectral face sensor grabbed a near-infrared and visible image spontaneously. The detector could simply detect the color and texture information of the counterfeit [7]. However, these texture analyses are constantly hampered by the unpractical condition and intricate algorithm.

Another advanced liveness detection technique is 3D Recognition, which postulated the face should have a depth feature if the face is a real one. However, this defending technique failed as the adversaries enhance their tactic. There was a team who build virtual 3D modal on the generic mobile phone screen and easily attack the ordinary system [25]. The virtual reality system had been augmented to a sufficient level to fortify any kind of user to bypass. Traditional 3D detecting measure was mainly depend on optical flow analysis, the motion speed of the central is higher than the outer region [8]. Some generated the optical flow from the whole face rather than part of it [9]. Other parts of the face suggested a liveness detection algorithm which analyzes the optical flow in detecting ears, nose, and mouth [10]. This algorithm is not effectively on depth character detection. Recently, the light field camera was contrived to optimize this mechanism. Through the direction and the intensity of the incoming light rays, the camera pictured the details of depth information just on a single image [11]. However, it is not pragmatic to generate special light conditions for these solutions.

Although some of these advanced liveness detection techniques like texture-based approach and 3D recognition work well against presentation attacks, they have requirements of special hardware like high-resolution camera and light field camera. For consumer electronics, people would like to pay more attention to the cost rather than security, and it is good enough to have the ability to combat photo-based attacks and video-based attacks. A liveness detection technique that does not require the addition of special peripherals is more suitable for mobile phones and other consumer electronics.

In this paper, our contributions are mainly included:

1) We introduced the popular presentation attacks and presentation attack detections on face recognition technology.

2) We conducted a study to propose a face liveness detection mechanism based on corneal reflection for smartphone.

3) We performed a series of experiments to evaluate the practicality, robustness and usability of CorneaReflect, respectively.

Experimental results show that CorneaReflect can effectively detect photo-based attacks and video-based attacks, and face authentication in different situations also has a considerable success rate.

1 PRELIMINARIES

1.1 FaceAuthentication for Smartphone

Face authentication for smartphones generally consists of two subsystems: one is face recognition and one other is face liveness detection, as Fig. 1 shown below. The face recognition subsystem is responsible for confirming whether it is a legitimate user. The face liveness detection subsystem is responsible for confirming whether the verification object is a living body. Only when both subsystems are authenticated, the face authentication system determines that the identity verification is successful. The face recognition subsystem obtains the face data of the verified user through the front camera of the mobile phone, and compares it with the face data of the legitimate user in the database. When they are consistent, it is considered a legitimate user. The face liveness detection subsystem considers the verification object to be a living body by detecting that the verification object has a certain specific living characteristic through the front camera or other sensors of the mobile phone. The face detection subsystem is an important module for mobile phones against presentation attacks. Different sensors or cameras are used to achieve different face liveness detection mechanisms.

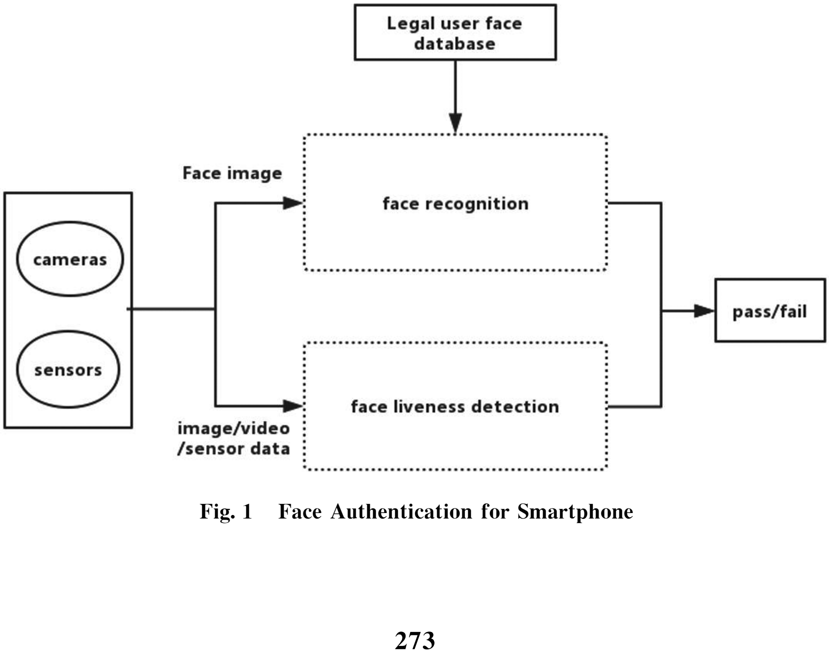


Fig. 1 Face Authentication for Smartphone

1.2 Presentation Attack and The Problem to Be Solved

Face authentication is to capture a facial image through a camera to identify a legitimate user, so the attackers spoof the face authentication system by presenting a fake face of a legitimate user in front of the camera. This is called presentation attack [25]. Presentation attack poses a serious threat to the face authentication system. The attacker uses photos, videos, or 3D masks to disguise a legitimate user, and the face recognition subsystem is unable to distinguish between the face biometrics taken from a living person and the face biometrics forged from the person's facial photos/videos/3D mask. That is why the face authentication system introduces a face liveness detection [28] subsystem to combat presentation attack.

Face liveness detection depends mainly on the information obtained by the sensor or camera. There are many existing face detection mechanisms, including challenge response approach, hybrid approach, texture-based approach and 3D recognition approach. It is very effective against 2D forgery attacks by the 3D recognition approach, but it requires the light field camera or other special hardware. Texture-based approach also has requirements of high-resolution camera. Both of these approaches add hardware cost, making it a low-cost solution for smartphones. Although the challenge response method does not require the addition of special hardware, it has a long interaction time and a poor user experience. Therefore, a practical and feasible solution should be proposed to solve all these problems.

In fact, making a 3D mask requires high-resolution face photo from multiple angles, which is very difficult and costly. For smartphones, it is good enough to have the ability to combat photo-based attacks and video-based attacks. In this paper, our proposed face liveness detection mechanism, CorneaReflect, aims to prevent both photo-based attacks and video-based attacks without adding hardware and interaction.

1.3 Corneal Reflection

Cornea is a transparent substance covering the pupil and iris of the human eye, which has a strong refractive ability. Its surface is covered with a thin layer of tear, which helps to moisten and smooth the cornea, to make the surface of the cornea have mirror-like reflection characteristics [13]. Light is reflected not once but four times from the eyes, although some are hard to see. These reflections are named Purkinje images [14], as Fig. 1 shown below. The reflection on the outer surface of the cornea is the strongest [15].

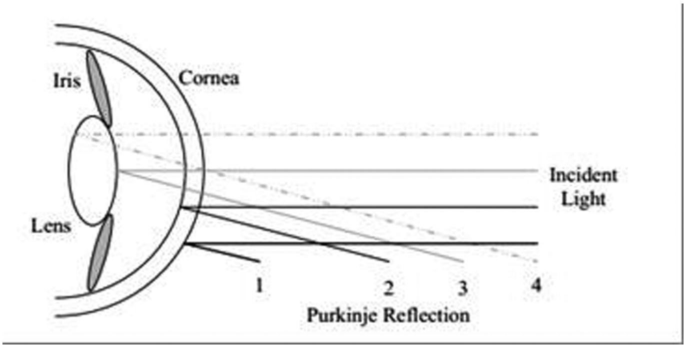


Fig. 2 Purkinje Reflection

When people stare at a light source, the reflection in their eyes will fall on the area of the pupil, which is generally black [15]. This makes it easy to distinguish the corneal reflection from the black pupil. When you take a picture of a person's face with some higher-resolution cameras, you can see clearly what's reflected by looking at the eyes.

Corneal reflection has been used in some research fields. For example, the most popular 2D function interpolation method for establishing remote gaze points (RGE) was achieved by using the vector between pupil center and the corneal reflection [16]. Michael Backes et al. extracted the clear letter image information reflected by the cornea from the human eye image captured by the magnifying camera lens [17]. Rob Jenkins and Christie Kerr use a high-resolution camera to take pictures of people to identify other people's faces through corneal reflections in order to provide a way to confirm the identity of the suspect[18].

Research mentioned above shows that the real cornea is reflective. By contrary, the cornea of the face in the general photos and videos is not reflective, which will be proved in Section 3. 1.

2 OUR APPROACH

In this paper, we proposed CorneaReflect, a practical and robust face liveness detection mechanism for smartphone based on the corneal reflection to counter the photo-based and video-based attacks. The mechanism is shown in the Fig. 3 below. First, the user needs to place the face in front of the front camera on the smartphone. Then, CorneaReflect allows the smartphone screen to display some random information and reflect it by the cornea. During the face authentication, the front camera of the smartphone captures the reflection of the smartphone screen on the cornea. As we know, the real cornea is reflective while the cornea of the face in the general photos and videos is not reflective. If the reflection and display are the same, the cornea is real, otherwise it is a forgery. In this way, CorneaReflect can effectively tell whether this is a photo-based/video-based attack or not without additional hardware and complex user interaction.

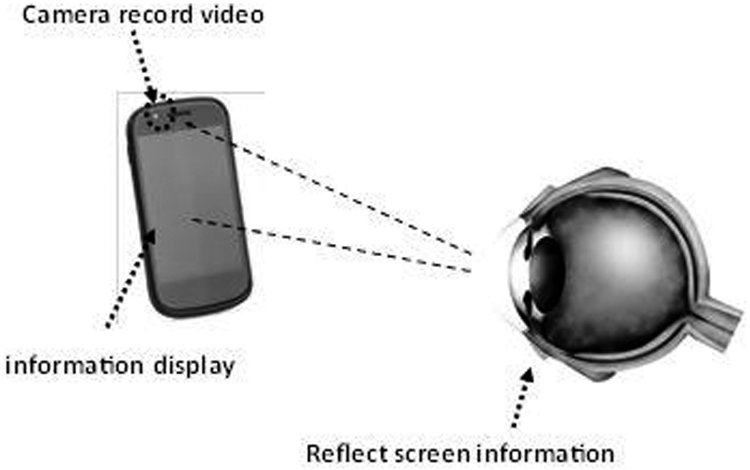


Fig. 3 The Liveness Detection Mechanism for Smartphone Based on Corneal Reflection

2. 1 Design overview

CorneaReflect consist of three modules, as shown in Fig. 4. First, the video capture module displays the information on the phone screen and calls the front camera to record the video of the entire face authentication process. Then, the video analysis module analyzes the face authentication video and extracts the information from the video. Finally, the results analysis module compares the information extracted from the video with the information from the video capture module. If they are the same, the corneal is real, otherwise it is forged.

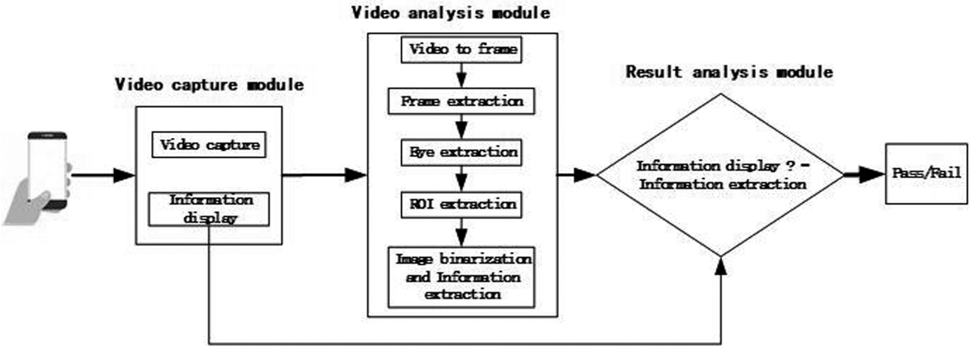


Fig. 4 The system view of CorneaReflect

2.2 Information displayed on smartphone screen

The information displayed on smartphone screen can be photos, videos, a color sequence, etc. What kind of information to display needs to consider two important factors. One factor is to ensure that information can be captured and identified correctly by the camera. In fact, light intensity, camera resolution, ambient illumination and other related factors have a great impact on the correct identification of information, especially when the information is photos and videos with lots of details. Daniel f. Smith found that in the case of strong light, even colors are difficult to recognize. And finally he had to conduct the experiment in a dark house [44]. Our design is to randomly display several black or white photos one by one on the screen of the smartphone. These black or white pictures consist of a black-and-white photo sequence and are much easier to be identified correctly than photos and videos with many details.

The other factor is to ensure that the mechanism can counter photo-based attacks and video-based attacks. If the attacker does not know the black-and-white photo sequence displayed on the smartphone screen, then the correct face verification video cannot be forged. Therefore, the black-and-white photo sequence should be as random as possible, making it difficult for attackers to guess. Assuming that n black or white pictures are displayed, there are only 2 choices for the photo, then there are 2^n sequences in total. If the sequence is treated as a password, the size of the entire password space is 2ˆn. To make the password space large enough to combat attacks, n needs to be as large as possible. But the larger n is and the more pictures are displayed, the longer the time for authentication will become. Therefore, the value of n needs to be balanced.

Our design allows the smartphone screen to display 10 photos, so the password space reaches $2^1 0$, which is 1024 size and has high security and ensures that the verification time will not be too long. Assume that the black photo indicates '0' and the white photo indicates '1'. If the random information displayed the smartphone screen is 0000001111, then the smartphone screen will display the photos of "black, black, black, black, white, white and white" one by one. As shown in Fig. 5.

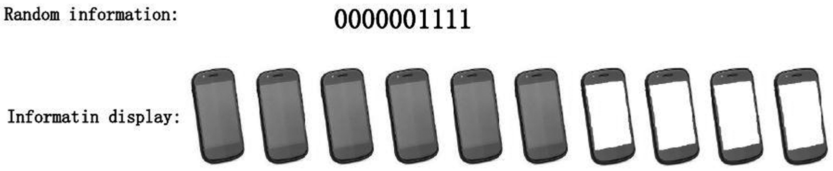


Fig. 5 Random information to Information display

When the black-and-white photo sequence is displayed too quickly, it may not be able to distinguish each photo. The photos displayed on the screen of the smartphone are equivalent to a continuous signal, as shown in Fig. 6. Camera records video, which is equivalent to sampling the continuous signal displayed on the screen. Video frame rate ***fps*** is equivalent to the sampling frequency ***fs. max***, and photo switching frequency ***sps*** of the screen is equivalent to the frequency of the continuous signal ***fmax***. According to the sampling theorem [22], if

$$\boldsymbol{fs.max} > \boldsymbol{2 * fmax} \qquad (1)$$

Then the continuous signal could be recovered. Therefore, if

$$\boldsymbol{fps} > \boldsymbol{2 * sps} \qquad (2)$$

Each photo of the black-and-white photo sequence coulddistinguished.

In our design, the ***fps*** is between 10 and 30 p/s, and ***sps*** is 2 p/s.

2.3 Information extraction and analysis

After the process of face authentication is completed, the video capture module sends the video of the face authentication process to the video analysis module. The video analysis module needs to perform a series of processing to extract information from the video, as shown in Fig. 4. The related processing is as follows:

(1)Convert video to frames.

Video needs to be converted to frames for further processing to get information. After the conversion, a video becomes several frames.

(2)Frame extraction.

Information will be extracted from the frames which converts from the video. Some of the frames contains the same information. For example, assuming the video frame rate is 10p/s and each photo is displayed on the screen for 1 second, then the captured video will have 10 frames for each photo. Therefore, it is enough to select one frame for each photo to analyze the information.

But it is not arbitrary to select a frame for analysis. When each photo displayed on the smartphone screen switches to another photo, the switching process lasts for a short period of time, which is called the response time, usually in the millisecond level. During this response time, the screen of the smartphone is switched between black and white photos. This intermediate state can be captured during camera recording, which is called a semi-illuminated frame, as shown in Fig. 6. The semi-illuminated frame could result in an inability to confirm whether the photo displayed on the screen is black or white. Therefore, when selecting frames for analysis, try to avoid selecting the semi-illuminated frame.

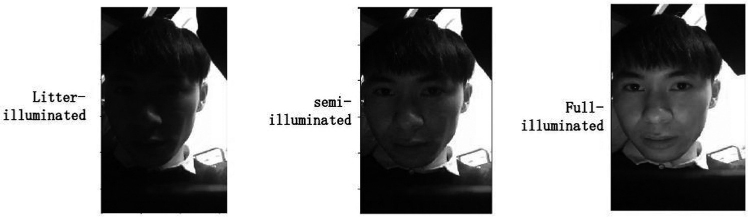


Fig. 6　semi-illuminated frame

The best way to invalidate a semi-illuminated frame is to select a frame that is captured in the intermediate instant of each photo display period, which is called an intermediate frame. They are located at the most stable moments during each photo display, usually not a semi-illuminated frame. Assuming that the frame rate is ***fps*** and photo switching frequency is ***sps***. Then he intermediate frames of ***m*** th photo in the ***l*** th second should be:

$$p=l*fps+m*\left\lceil\frac{fps}{sps}\right\rceil+\left\lceil\frac{fps}{2sps}\right\rceil \tag{3}$$

Here ***p*** is the intermediate frame. All intermediate frames can be found by the above formula. When all the intermediate frames are found, the frame extraction is completed.

(3)Eye extraction.

After the intermediate frame extraction, the video has been converted into several facial images. We need to further extract the image of the eye. There have been plenty of researches on eye extraction. For example, Yuille used template to extract eye features [19]. Kampmann and Zhang et al. used luminance edges and the difference in brightness between eye white and iris to extract eye features [20], Feng Others use eyelids and eye corners to extract eye features [21]. Here we use face_recognition-an open source library for eye image extraction. It is based on the deep learning model in the industry-leading C++ open source library dlib. The accuracy of face recognition is as high as 99.38%. It can identify a person's face, eyes, eyebrows, mouth, etc. Fig. 7 shows the images after eye extraction.

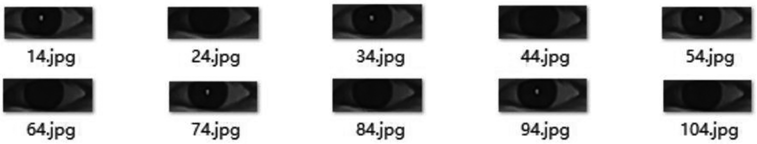


Fig. 7 eye extraction

(4)ROI extraction.

After eye extraction, it is necessary to further extract the image of the smartphone screen from the corneal reflection to get information. The reflection of the smartphone screen is a small image near the center of the pupil, which is the region of the interest, called the ROI. In order to extract ROI, the center of the pupil should be located first.

Hough transform is a good method to detect the pupil and locate the center of the pupil. However, there are many details in the eye image, which is the noise of the Hough transform and will interfere with pupil detection. Therefore, some image preprocessing needs to be performed first, including conversion to grayscale image, open operation, canny edge detection, in order to remove noise and reduce data amount.

First, the eye image is converted from the RGB image to the grayscale image in order to reduce the amount of data. The conversion formula is as follows:

$$\boldsymbol{Gray} = \boldsymbol{R} * \mathbf{0.299} + \boldsymbol{G} * \mathbf{0.587} + \boldsymbol{B} * \mathbf{0.114} \tag{4}$$

The RGB image is converted to the grayscale image, as shown in Fig. 8.

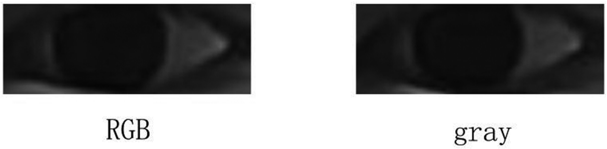


Fig. 8 RGB to Gray

Then, an open operation is used to eliminate noise and remove the details of the eye image. A 5 * 5 convolution kernel is used to perform the open operation in the eye image. After the open operation, the image changes are shown in Fig. 9. After these processing, some details in the eye image are blurred and the pupil's outline becomes more distinct.

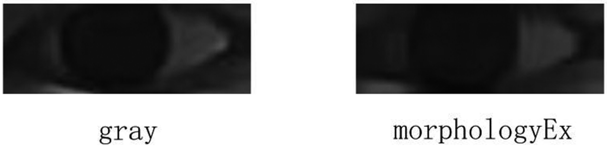


Fig. 9 Fig. 8. RGB to Gray

In order to extract the pupil image better, the canny edge detection is used to extract the edges, remove the extra details, and greatly reduce the amount of data. Canny edge detection is a very popular edge detection algorithm proposed by John F. Canny in 1986 [22]. The algorithm first converts the image into a grayscale image, and then uses a 5 * 5 Gaussian filter to smooth the noise. The Gaussian function is as follows:

$$\boldsymbol{H}(\boldsymbol{x},\boldsymbol{y}) = \frac{\mathbf{1}}{\mathbf{2}\boldsymbol{\pi}\boldsymbol{\sigma}^2}\mathbf{exp}\left(-\frac{\boldsymbol{x}^2+\boldsymbol{y}^2}{\mathbf{2}\,\boldsymbol{\sigma}^2}\right) \tag{5}$$

we calculate the first derivatives (G_x and G_y) of the horizontal and vertical directions in the Gaussian smoothed image, and according to the obtained two gradient maps (G_x and G_y) we can find the gradient and direction of the boundary. The formula is as follows:

$$\boldsymbol{Edge\_Gradient}(\boldsymbol{G}) = \sqrt{\boldsymbol{G_x}^2 + \boldsymbol{G_y}^2} \tag{6}$$

$$\boldsymbol{Angle}(\boldsymbol{\theta}) = \mathbf{tan}^{-1}\left(\frac{\boldsymbol{G_x}}{\boldsymbol{G_y}}\right) \tag{7}$$

After obtaining the gradient and direction, the entire image is scanned, and each pixel is examined, and the point, which gradient is the largest among the points having the same gradient direction, was selected as candidate point of the edge. Finally, the double-threshold method is used to find out all the candidate points and stitch them into contours: for a candidate point, if its gradient is greater than the high threshold, it is used as the edge point; if it is lower than the low threshold, it is discarded; if it is between the low threshold and high threshold, only when the point is connected with other edge point, it is used as the edge point.

After the above pre-processing, the eye image only retains some contours, and then using the Hough transform to extract the pupil image becomes very reliable. Because the pupil is circular and often obscured by the eyelids, this case can segment the pupil well by the Hough transform [22]. Using the Hough transform to detect a circle, the fundamental is that any point $\{(x_i, y_i) | i=1,2 \cdots n\}$ in the image could be part of the candidate circle set [23]. Knowing that a point (x_i, y_i) is on the circumference, we need to find out the circle radius r and the center coordinates (a, b), the following formula is obtained:

$$(\boldsymbol{a}-\boldsymbol{x}_i)^2+(\boldsymbol{b}-\boldsymbol{y}_i)^2=\boldsymbol{r}^2 \tag{8}$$

The point in the (x, y) plane is transformed into a three-dimensional cone on the parameter space (a, b, r), and the point on the cone corresponds to the circle of all the points (x_i, y_i) on the original image. If all non-zero pixels in the (x, y) plane are converted to these point sets in the (a, b, r) space and their contributions are added, then the point where the local maximum occurs on the accumulated space (a, b, r) corresponds to the circle (a_0, b_0, r_0) where the original image appears on the (x, y) plane.

Using the Hough transform, the center of the pupil (x_0, y_0) can be found. As shown in Fig. 10, the white point is the center of the pupil, and the white circle is the edge of the pupil.

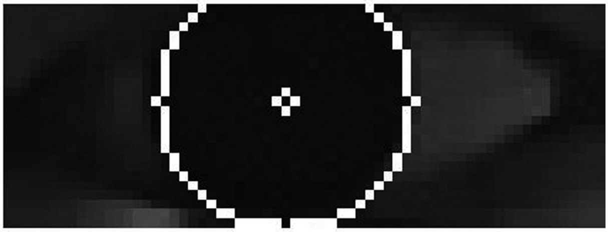


Fig. 10 Hough transfrom

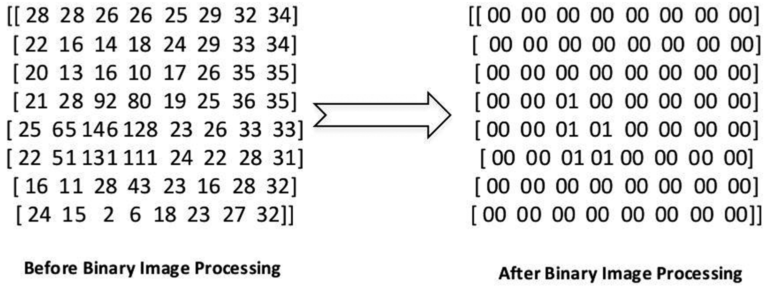
Therefore, ROI can be extracted from the center of the pupil.

(5)Information extraction.

After a series of image processing, the eye images have become the ROI. The information can be extracted from the ROI.

The color depth between white and black is divided into several levels, called 'grayscale'. The range is generally from 0 to 255, while the white is 255 and the black is 0. While the smartphone screen displays a white photo, the ROI has pixels with a larger gray scale value. While the smartphone screen displays a black photo, the gray scale value of all pixels in ROI will be smaller. Experiments show that if the ROI has a pixel with a gray value greater than 90, it means that the smartphone screen displays a white photo, otherwise it is a black photo.

In order to extract information, the gray value is binarized. If the gray value is greater than or equal to 90, the gray value is set to 1; And if it is less than 90, the gray value is set to 0. As shown in Fig. 11.



```
[[ 28  28  26  26  25  29  32  34]
 [ 22  16  14  18  24  29  33  34]
 [ 20  13  16  10  17  26  35  35]
 [ 21  28  92  80  19  25  36  35]
 [ 25  65 146 128  23  26  33  33]
 [ 22  51 131 111  24  22  28  31]
 [ 16  11  28  43  23  16  28  32]
 [ 24  15   2   6  18  23  27  32]]
```

Before Binary Image Processing

```
[[00 00 00 00 00 00 00 00]
 [ 00 00 00 00 00 00 00 00]
 [00 00 00 00 00 00 00 00]
 [00 00 01 00 00 00 00 00]
 [00 00 01 01 00 00 00 00]
 [00 00 01 01 00 00 00 00]
 [00 00 00 00 00 00 00 00]
 [00 00 00 00 00 00 00 00]]
```

After Binary Image Processing

Fig. 11 Gray Value Binarization

Then, if there is 1 in the pixel matrix of ROI, it represents that the smartphone screen displays a white photo, otherwise it represents that a black photo is displayed.

At the end of the analysis, a black-and-white photo sequence can be obtained. The analysis result is compared with the known information displayed on smartphone screen. If they are consistent, the face authentication succeeds, otherwise it fails.

3 EVALUATION

The study is divided into three parts to evaluate the practicality, robustness and usability of CorneaReflect, respectively. In this section, the experimental process, analysis process and the experimental results will be described in detail.

3.1 Practicality experiment

As mentioned before, research indicates that the real cornea is reflective. But no research shows that the cornea in the photo and video is not reflective. Therefore,an experiment was carried out to prove this. However, the attacker may deceive the system by placing a plane mirror or convex mirror on the cornea in photos and videos. Another experiment was conducted to determine whether it works or not.

1)Experiment setup

Experiment 1:The experiment selected 100 different 4032 * 3016 pixels HD face photos. These face photos were printed as paper photos using 12-inch glossy photo paper. In addition, these face photos are close to the real head size displayed on a 32-inch 4K display. Paper photos and monitors are placed in front of the front camera of the phone, pretending to be a real person for face verification. The video of the face authentication process will be converted into frames to analyze whether the cornea can reflect the information on the phone screen. If it can, record it as “yes”, otherwise record it as “no”.

Experiment 2: The experiment places a plane mirror and convex mirror on the cornea in photos and videos. The photos and display are the same as experiment 1. If the attacker can deceive the system in this way, record it as “yes”, otherwise record it as “no”.

2)Experiment result

Experiment 1: In the experiment, a large number of frames need to be analyzed. So, we wrote a small program to see if the cornea of photos and videos have reflections on the screen of the phone. The result is shown in the Fig. 12 above. All the cornea both in photos and videos cannot reflect the information on the phone screen. Therefore, the cornea in photos and videos is non-reflective.

Experiment 2: The statistical results of the experiment are shown in the

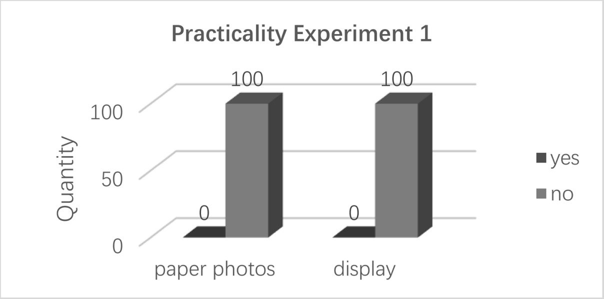


Fig. 12　Practicality experiment 1

Fig. 13 above. No attack succeeded, no matter whether a plane mirror or convex mirror on the cornea. Actually, there are very different reflection properties between the cornea and the plane mirror/convex mirror. Perhaps, by customizing an artificial forgery similar to the cornea, it is possible to successfully deceive the CorneaReflect.

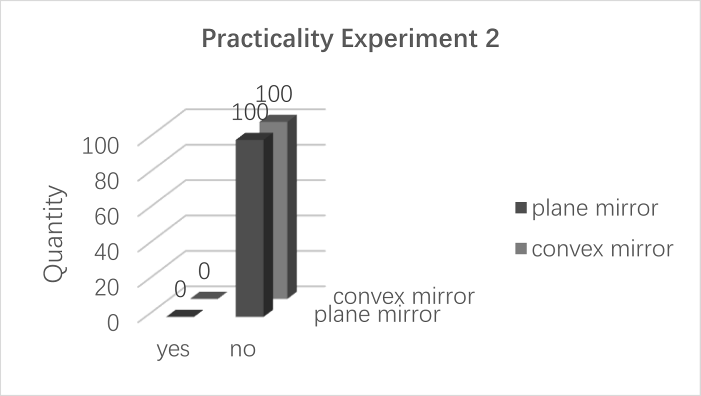


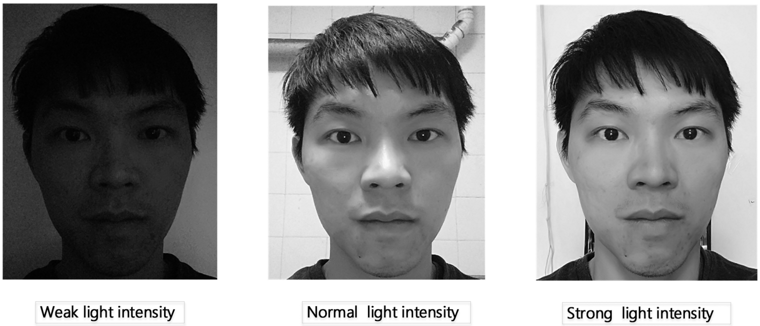
Fig. 13　Practicality experiment 2

3. 2　Robustness experiment

Robustness is very important for a system. We will verify the robustness of the CorneaReflect from three aspects: environmental factors, device factors and user factors.

1)Experiment setup

The experiment selected 100 participates. Among them, there are 50 males and 50 females, aged between 18 and 60, 25 injunior school or below, 26 in high school education, 30 in undergraduate degree, and 19 in master's degree. There are 85 yellow people, 7 black people, and 8 white people. The experiment will be performed under different conditions, including different light intensity (see Fig. 14) and ambient illumination. Besides, the experiment will be carried out on different smartphones with different screen size and camera resolution.



Weak light intensity

Normal light intensity

Strong light intensity

Fig. 14 different light intensity

All participants followed the experiment instructions below: Firstthey are asked to stay still, not wearing glasses. Then, they open the app in the smartphone and keep the distance between the phone and the face between 30 and 45 cm, so that the camera can capture the entire face and level with the eyes. Finally, click on the "Start" button to start the face authentication and record a video. During the face authentication process, the participants are required to keep their eyes open and look at the front camera. All participants are asked to complete the experiments under different light intensity, different ambient illumination, different screen size and different camera resolution.

The number of successes will be recorded and the success rate will be calculated.

2)Experimentresult

Environmental factors:

In the daily life, smartphone will be used under different environment, like different light intensity and many kinds of ambient illumination including lamp, computer screen, the light through windows, etc. They may affect the success of face authentication.

We count the success rate of all participants in the case of different light intensity and ambient illumination. The data is shown in Fig. 15. It shows that the intensity of light has a great impact on the success rate. A high success rate close to 100% will be achieved under a proper intensity of light. Under a strong intensity of light, success rate drops down to 66%. This is because the area of the smartphone screen reflected on the cornea under strong light intensity will become smaller than the area under normal light intensity. If the location of pupil center is not accurate, the ROI may not involve smartphone screen reflection, and that will result in an error of in information extraction. In the case of weak light intensity, the face will not be so obvious that the program will make mistakes in identifying the pupil center, which will also lead to information extraction errors.

Besides, Fig. 14 shows that ambient illumination has little effect on success rate. This is because when the user sees the front camera of the smartphone, the reflection of the ambient illumination in the cornea is always outside the center of the cornea.

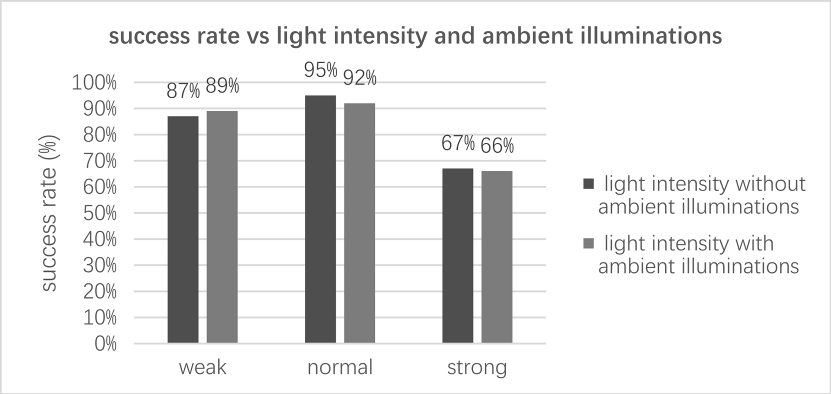


Fig. 15 Success rate vs light intensity and ambient illuminations

Device factors:

This experiment was conducted to reveal whether screen size and camera resolution have an impact on success rate. Xiaomi 6 mobile (5. 15 inch) and Xiaomi 8 mobile (6. 26 inch) were used to carry out this experiment. The front camera resolution can be set to 480 * 640, 720 * 1280, 1080 * 1920 three resolutions. We calculate the success rate of using different phone screen sizes and camera resolutions in normal light intensity and without ambient lighting. The result is shown in Fig. 16. Obviously, the screen size and camera resolution have little effect on the success rate. This is because CorneaReflect only needs to capture black/white photos and extract a small area of ROI for information extraction.

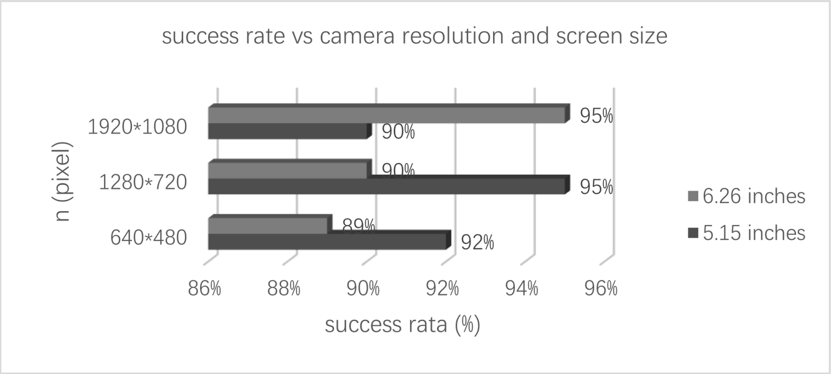


Fig. 16 Success rate vs camera resolution and screen size

User factors:

Face authentication is operated by the user, so the user must have a great influence on the CorneaReflect. In this experiment, we consider the impact of age, gender, education and ethnicity on CorneaReflect. Fig. 17 shows that user

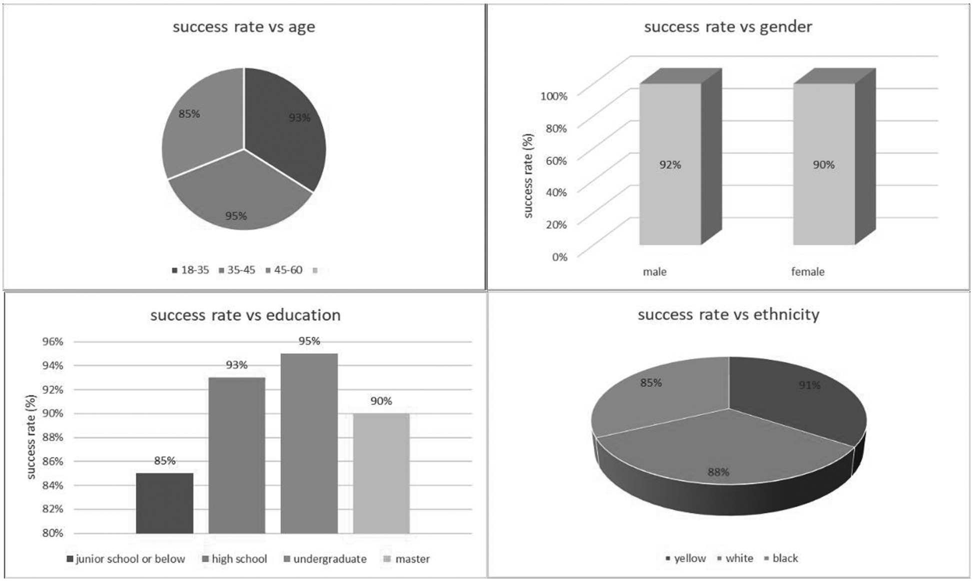
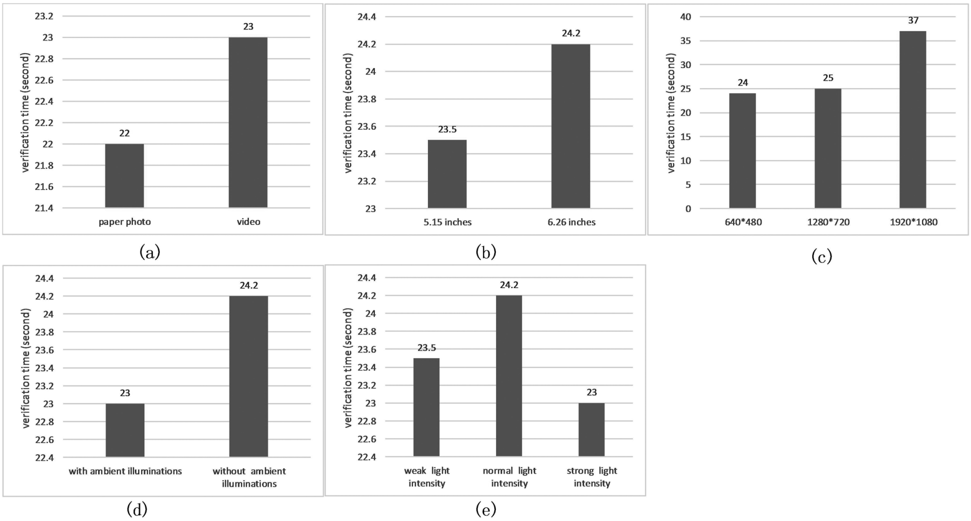


Fig. 17 Success rate vs user factors

factors have a litter influence on success rate. Participants between the ages of 45 and 60 have a significantly lower success rate than other age groups, only 85%. The success rate of participants between the ages of 18 and 35 is very similar between 35 and 45. Besides, the success rate of male and female are almost the same. Participants with junior school or below have a significant low success rate, and most of them are between 45 and 60 years old. Finally, we can see that there are some differences in the success rates between different races. But the number of some races is too small to be statistically significant.

3.3 Usability experiment

Verification time is a very important indicator of usability. In this experiment, we calculated the verification time of all the experiments conducted before. The result is shown in Fig. 18.



(a) (b) (c)

(d) (e)

Fig. 18 Verification time

It takes some time forCorneaReflect to send the videos of face authentication to the server for analysis, which we define as t_s. And t_s is determined by internet speed and the video size. Besides, it takes a longer time to analyse the video of face authentication, and we define the time as t_a. The t_a is determined by the video size. What's more, Smartphone screen will display black-and-white photo sequence, and we define the time as t_d. So, the verification time,

$$t_v = t_s + t_a + t_d \tag{9}$$

Fig. 18 shows the verification time of the five experiments. We can see that the verification time is only related to camera resolution. Verification time consists of internet speed and video file size. The higher the camera resolution is, the bigger the video file size is. It results in that t_s*and* t_a become bigger. Therefore, t_v will become bigger.

In our design, the verification time is too long. We should find out some method to solve this problem. In order to reduce the verification time, we should use 480 * 640 as the camera resolution. What's more, analyzing the video of face authentication on smartphone can delete the t_s, which will reduce the verification time a lot.

3.4 Experimental summary

The experiments prove the practicality and robustness of CorneaReflect. In the case of different environments, different devices and different users, the success rate is still high, the average value is about 90%. The experimental results show that the long verification time leads to poor usability and still needs to be improved.

4 CONCLUSION

In this paper, we propose CorneaReflect, a practical and robust face liveness detection mechanism for smartphone to counter the photo-based and video-based attacks based on the corneal reflection. The smartphone screen displays a black-and-white photo sequence which is reflected by the cornea. Then reflection on the cornea is captured by the smartphone's front camera. Finally, we can tell whether the reflected information matches the displayed information. The experiments confirmed that CorneaReflect is feasible, robust and usable, and it can effectively counter the photo-based and video-based attack.

References

［1］ L. O'Gorman. Comparing passwords, tokens, and biometrics for user authentication. Proceedings of the IEEE,91(12):2021-2040, 2003.

［2］ R. Ghiass, O. Arandjelovic, H. Bendada, and X. Maldague. Infrared face recognition: A literature review. In IJCNN 2013, pages 1-10, 2013.

［3］ J. Wilder, P. J. Phillips, C. Jiang, and S. Wiener. Comparison of visible and infra-red imagery for face recognition. In FG 1996, pages 182-187. IEEE, 1996.

［4］ Yan Li, Yingjiu Li, Qiang Yan, et al. Deng Seeing Your Face Is Not Enough: An Inertial Sensor-Based Liveness Detection for Face Authentication 2015

［5］ S. Marcel, M. S. Nixon, and S. Z. Li, eds. , Handbook of Biometric Anti-spoofing, Springer, 2014.

［6］ G. Fadda, M. Pili, N. Sirena, G. Murgia, M. Ristori, et al. Competition on counter measures to 2-d facial spoofing attacks. In IJCB 2011, pages 1-6. IEEE, 2011.

［7］ D. Yi, Z. Lei, Z. Zhang, and S. Z. Li. 2014. Face anti-spoofing: Multispectral approach. In Handbook of Biometric Anti-Spoofing, S ebastien Marcel, Mark S. Nixon, and Stan Z. Li (Eds.). Springer London, 83-102.

［8］ O. Kahm and N. Damer. 2d face liveness detection: An overview. In BIOSIG 2012, pages 1-12, 2012.

［9］ W. Bao, H. Li, N. Li, and W. Jiang. A liveness detection method for face recognition based on optical flow field. In IASP 2009, pages 233-236. IEEE, 2009.

［10］ K. Kollreider, H. Fronthaler, and J. Bigun. Non-intrusive liveness detection by face images. Image and Vision Computing, 27 (3): 233-244, 2009.

［11］ R. Raghavendra, K. Raja, and C. Busch. 2015. Presentation attack detection for face recognition using light field camera. IEEE Transactions on Image Processing 24, 3 (2015), 1-16.

［12］ G. Balakrishnan, F. Durand, and J. Guttag. Detecting pulse from head motions in video. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pages 3430-3437,2013.

[13] P. Gang, S. Lin, W. Zhaohui, and L. Shihong. 2007. Eyeblink-based anti-spoofing in face recognition from a generic webcamera. In IEEE 11th International Conference on Computer Vision, 2007 (ICCV'07). 1-8.

[14] Nitschke C, Nakazawa A, Takemura H. Corneal Imaging Revisited: An Overview of Corneal Reflection Analysis and Applications [J]. Ipsj Transactions on Computer Vision \& Applications, 2013, 5:1-18.

[15] Lee E C, Park K R, Kim J. Fake iris detection by using purkinje image [C]//International Conference on Biometrics. Springer, Berlin, Heidelberg, 2006: 397-403.

[16] Smith D F. Countering digital replay attacks for face verification on consumer smart devices using structured illumination[J]. 2016.

[17] Dan W H, Ji Q. In the Eye of the Beholder: A Survey of Models for Eyes and Gaze[J]. IEEE Transactions on Pattern Analysis \& Machine Intelligence, 2010, 32(3):478.

[18] Backes M, Unruh D. Compromising Reflections-or-How to Read LCD Monitors around the Corner[C]// IEEE Symposium on Security and Privacy. IEEE Computer Society, 2008:158-169.

[19] Yuille A L, Hallinan P W, Cohen D S. Feature extraction from faces using deformable templates[J]. International journal of computer vision, 1992, 8(2): 99-111.

[20] Kampmann M, Zhang L. Estimation of eye, eyebrow and nose features in videophone sequences[C]//International workshop on very low bitrate video coding (VLBV 98). 1998: 101-104.

[21] Feng G C, Yuen P C. Multi-cues eye detection on gray intensity image [J]. Pattern recognition, 2001, 34(5): 1033-1046.

[22] Canny J. A computational approach to edge detection[M]//Readings in computer vision. Morgan Kaufmann, 1987: 184-203.

[23] Sonka M, Hlavac V, Boyle R. Image processing, analysis, and machine vision[M]. Cengage Learning, 2014.

[24] Ballard D H. Generalizing the Hough transform to detect arbitrary shapes[J]. Pattern Recognition, 1981, 13(2):111-122.

[25] Ramachandra R, Busch C. Presentation attack detection methods for face

recognition systems: a comprehensive survey[J]. ACM Computing Surveys (CSUR), 2017, 50(1): 8.

[26] Yi Xu, True Price, Jan-Michael Frahm, and Fabian Monrose, Virtual U: Defeating Face Liveness Detection by Building Virtual Models from Your Public Photos August 10-12, 2016.

[27] R. K. Rowe, U. Uludag, M. Demirkus, S. Parthasaradhi, and A. K. Jain. A multispectral whole-hand biometric authentication system. In Biometrics Symposium, 2007, pages 1-6. IEEE, 2007.

[28] Tan X, Li Y, Liu J, et al. Face liveness detection from a single image with sparse low rank bilinear discriminative model[C]//European Conference on Computer Vision. Springer, Berlin, Heidelberg, 2010: 504-517.