単眼カメラでの顔認証なりすまし防止

Face Spoof Detection Method for Monocular Camera

曹　永剛*　　顧　炯*
Yonggang CAO　　Jiong GU

要　旨

顔認証の技術は今急速に発展している．セキュリティ監視システムやスマートフォンだけでなく，オフィスプリンティング分野でも製品に顔認証を組み合わせることが試みられている．顔認証の応用の拡大に伴い，その安全性の確保が重要となっている．人の顔の画像や映像は入手しやすいため，なりすましに対する検査が必要である．本論文では，赤外線や構造化光を用いない安価な単眼カメラを使い，顔写真やビデオによるなりすましを検出する低コストのシステムを提案する．顔検知，輝度確認，鮮明度確認を行った後，RGB，HSV，CNNの3つの手法でなりすまし検出を行い，それらを統合した分類器で最終結果を得る．統合により，単一手法より高いロバスト性が得られた．

ABSTRACT

Facial recognition technology has been developing rapidly nowadays for applications such as mobile phones, access controls, and security monitoring systems. Furthermore, the feasibility of combining office automation products with face recognition is being explored. The security risks of facial recognition are becoming exposed as more applications emerge. Because faces are frequently exposed in photos and videos, it is necessary to implement face spoof detection. Our solution introduces a low-cost face spoof detection system, which enables face photo and video spoof detection using inexpensive monocular cameras (compared with infrared cameras and 3D structured light cameras). In the pretreatment stage, face detection, brightness evaluation, and image blur detection are used. In the core processing stage, RGB, HSV, and CNN methods are used to detect spoof faces. Finally, an integrated classifier is designed based on the output of the above three classifiers, making it more robust than a single system.

* 理光図像技術（上海）有限公司
Ricoh Imaging Technology (Shanghai) Co., Ltd.

1. Background and Target

RITS (Shanghai) developed the RiTrac printing system based on Multi-Functional Peripheral (MFP), which achieved successful sales in the Chinese market. However, we noticed customers prefer face login over smart cards because smart cards can be lost and are expensive to replace. We intend to add a face login function to the RiTrac printing system. Because the MFP smart panel only carries the basic UVC camera driver, it can only be used with a monocular camera. Therefore, a face spoof detection system suitable for monocular cameras is needed to resolve security problems that arise in practical application.

At present, spoof face detection based on a monocular camera requires user action. For example, the user may be asked to shake his or her head[1]. Alternatively, the user's eye movement[2] (blinking) may be detected by the optical flow method. However, both methods are not failproof, and they often cannot prevent face video attacks as actions such as head shaking and blinking can be easily recorded and replayed. Current solutions involve infrared, laser, and multi-camera technology, though many challenges still persist.

Fig. 1 illustrates the architecture of our facial recognition system, which is deployed in the on-premise mode. The MFP obtains the face image via the USB camera and sends it to the server in the local area network. The server then runs the spoof face detection and facial recognition program.

Fig. 1 Face login service scenario for RiTrac system.

2. Solution Overview

It is widely known that face spoof detection based on image texture is easily affected by environmental illumination. Meanwhile, face spoof detection based on neural networks is prone to over-fitting due to insufficient samples. Therefore, two methods are considered in our scheme: one is to add a pretreatment stage so that only face images with sufficient brightness and sharpness can enter the real detection stage; the second is to train an integrated classifier with the output (probability) of an image texture classifier and a neural network classifier.

It should be noted that the face area is detected first, and brightness and definition are evaluated afterwards because the object of brightness and definition detection must be the face area, not the background area.

The entire process, shown in Fig. 2, is divided into two stages. The first stage is pretreatment which includes the evaluation of the basic image quality (brightness and clarity) of the face area, and the second stage is the detection.

This process should be carried out before facial recognition. A real face cannot enter the next stage of facial recognition until passing this detection. After the user is identified, they can login to the system.

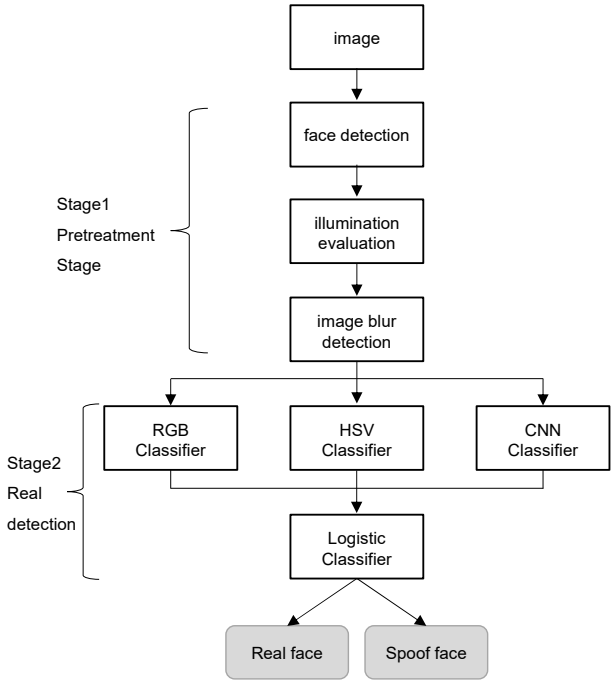

Fig. 2 Two-stage face spoof detection.

3. Pretreatment Stage

The purpose of pretreatment is to select face images that meet the requirements for ambient illumination and definition.

3-1 Face Detection

We used OpenCV Cascade Classifier (Face Detector) to detect whether an image contains a face. While it is not a state-of-the-art face detector, it is suitable for our work. Because face occlusion cannot pass this detector, we can obtain complete face images without any interference.

3-2 Illumination Evaluation

We use gray histograms to calculate the brightness of the images. We focus on whether the brightness of the face meets the requirements, so we need to crop the local image of the face, resize it to 64×64 pixels, and then convert it into a gray image as shown in Fig. 4. In the previous step, face detection returned to face contour. First, the gray scale value of each pixel in the face is counted (see Fig. 3) and then the mean value is calculated. By specifying the upper and lower limits of the mean value, we can ensure that the brightness of the face is in a suitable range.

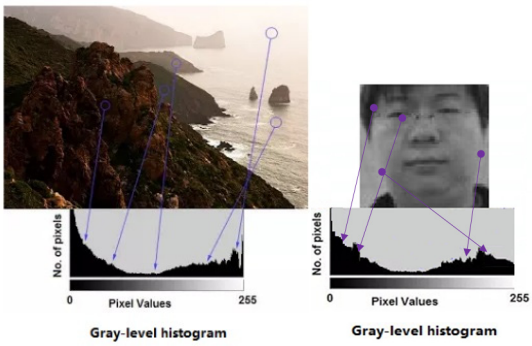

Fig. 3 Brightness distribution represented by histogram.

Fig. 4 64×64 pixels cropped face image, RGB to grayscale.

3-3 Image Blur Detection

Motion and unfocused cameras can lead to blurred photos. If the photos are too blurred, they are not suitable for subsequent color and texture detection. We use OpenCV Laplace operator to calculate the image blur score of the 64×64 face image. A threshold is needed to ensure that the image is not too blurred.

To determine the threshold, we investigated the face image samples, removed the outliers through box plot method, and determined the lower limit of blur.

4. Face Spoof Detection

Face spoof detection consists of the following three components: RGB image features, HSV image features, and neural network classification.

The face samples were collected from the company in which the experiment was conducted. A USB camera was used to continuously record video, and then the image frames were extracted from the footage. A total of 100,000 sample faces were collected from photos and mobile phones. Because the mobile phone display is very small, the phone needs to be close to the camera when collecting the face image.

- Sample collectors: 30 persons (15 males, 15 females)
- Total number of samples collected: 100,000
- Camera model: Logitech C925e
- Printing paper: A4 normal
- Mobile phone model: 10 phone models

Table 1 shows the sample composition (the value is the proportion).

Table 1 Composition of face samples.

	Real face	Printing photo face	Phone photo face	Phone video face
Office lighting (About 4500k)	25%	5%	5%	5%
Sunlight (day & window)	15%	5%	5%	5%
Sunlight (dusk & window)	15%	5%	5%	5%

4-1 RGB Image Features

The reflection and refraction of light causes noise in face images captured twice, e.g., on a mobile phone screen and a photograph, which makes the distribution of the RGB pixel histogram wider than that of a real face image.

Face images are segmented into three channels: R, G, and B. The histogram statistical features of each channel are extracted and normalized. For each histogram, the Y coordinate represents the total count at one grayscale; therefore, for 256 grayscales there are 256-wide counts on the X coordinate, i.e., we obtain the Y coordinate result as a 256-length vector. Concatenate these three RGB vectors into a single 768-length feature vector (see Table 2)

$$hist_{concatenate} = hist_R \cup hist_G \cup hist_B$$

and then train an Extra Trees Classifier.

Table 2 Sample RGB feature vector.

$hist_R$	$hist_G$	$hist_B$
0,0,4,23,…,32,0,0	0,0,4,89,…,32,0,0	0,0,4,89,…,32,0

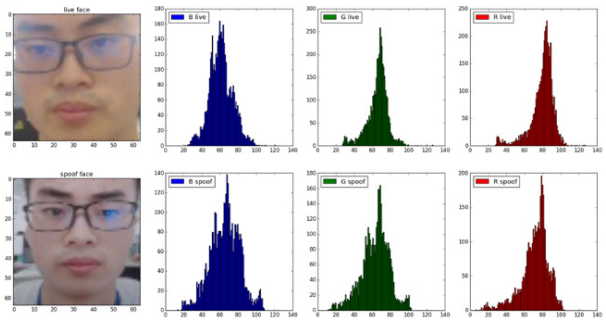

Fig. 5 RGB image features - The top image is a real person and the bottom image is a spoof face.

It is evident that the distribution of the RGB histogram is wider than that of the real face image (see Fig. 5).

4-2 HSV / HLS / LUV / LBP Image Features

The face images are converted from RGB to HSV (H chroma, S saturation, V brightness). The spatial distribution of the real faces differs from that of the spoof faces. We transformed 64×64 pixels face images into HSV, HLS, and LUV color space. We extracted the first-order moments (mean), second-order moments (variance), and third-order moments (slope) of each space one by one. We also extracted LBP features based on the grayscale of source image. We concatenated these features into a 1040-length long vector and then trained an Extra Trees classifier.

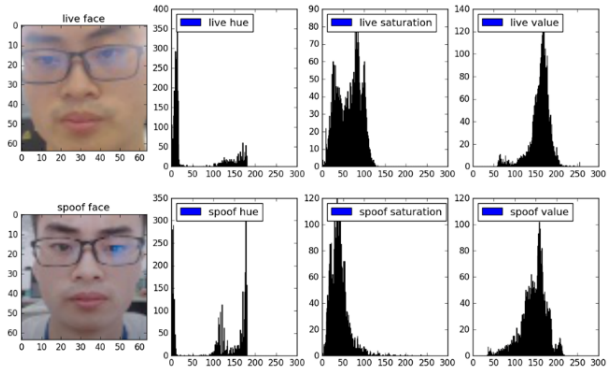

Fig. 6 HSV image features - The top image is a real person and the bottom image is a spoof face.

The spoof face saturation is relatively low due to specular reflection, as seen in Fig. 6.

The color spaces feature-extraction goes through every 16*16 window. For the 64×64 source image, we obtain 16 windows, and we perform LBP iterations three times with different (P, R) parameters, (P=8, R=1), (P=5, R=2), and (P=16, R=2). The algorithm is illustrated in Fig. 7.

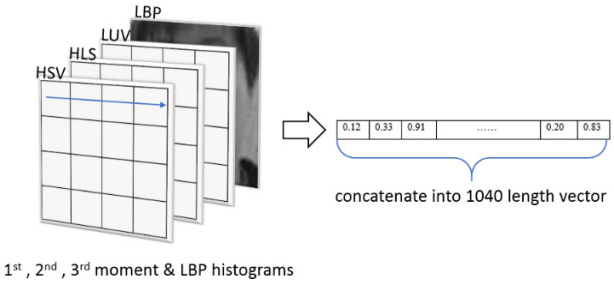

Fig. 7 Multiple color spaces & LBP image features concatenated into long vector (window size is 16*16 at each iteration).

4-3 Neural Network Classification

We used FasNet[3] (VGG16 Backend), a leading open source neural network classifier, to train a classifier with 100,000 real face and spoof face samples under different illumination conditions. Here, two classes, real faces and spoof faces, are used.

4-4 Integrated Classifier

Although HSV and RGB classifiers are easily affected by brightness, they perform steadily under suitable conditions. Meanwhile, neural network classifiers are less susceptible because samples under different illumination (sunlight, fluorescent lamp, etc.) are added during training. However, its result is unstable, so we use the output results (probability) of the three classifiers to train a logistic regression classifier as shown in Fig. 8 to attain more precise differentiation.

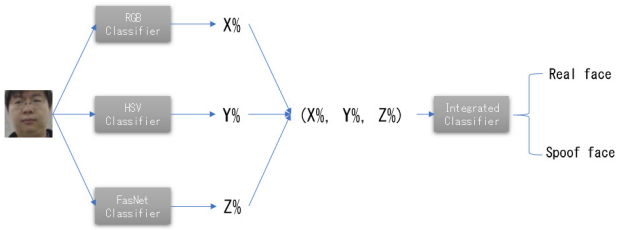

Fig. 8 Complete spoof face detection process.

As shown in Table 3, the rejection rate of the integrated classifier is higher than that of the base classifiers. The training set and the test set consist of 90,000 samples and 10,000 samples, respectively.

Table 3 Evaluation results of rejection rate.

RGB Classifier	HSV Classifier	FasNet Classifier	Integrated Classifier
95%	93%	96%	99%

5. Conclusion

In the company's internal test of our face spoof detection (under office lighting with 30 people and 10 types of mobile phones and printed photos), we attained a 97% pass rate (real person passing detection) and a 99% rejection rate (spoof face rejection rate). The above algorithm runs on the i7 CPU, GTX1070 graphics card computer and takes about 100 milliseconds. In the future, it will be applied to RITS (Shanghai) solution products.

Internal Test Conditions:

- Camera model: Logitech C925e
- Test subjects: 30 people (does not include sample collector)
- Printed photo faces: 30
- Phone photo faces: 30
- Phone video faces: 30

Test Method:

1. Pass rate – Real face, each person try 5 spoof face detection, counts the number of pass.
   * Pass rate = pass times / (30*5)
2. Rejection rate – Spoof face (printed photo, phone photo, phone video) is subjected to spoof face detection three times and the number of rejections is counted.
   * Reject rate = number of rejections / (30*3 + 30*3 + 30*3)

References

1) S. Kawato, J. Ohya: Real-time Detection of Nodding and Head-shaking by Directly Detecting and Tracking the "between-eyes," *Proceedings Fourth IEEE International Conference on Automatic Face and Gesture Recognition (Cat. No. PR00580) IEEE*, pp. 2–3 (2002).
2) G. Pan et al.: Eyeblink-based Anti-spoofing in Face Recognition from a Generic Webcamera, *IEEE 11th International Conference on Computer Vision, ICCV 2007*, pp. 1–2 (2007).
3) O. Lucena et al.: Transfer Learning Using Convolutional Neural Networks for Face Anti-spoofing, *International Conference Image Analysis and Recognition Springer, Cham*, pp. 3–5 (2017).