Liveness Detection for Embedded Face Recognition System

Hyung-Keun Jee, Sung-Uk Jung, and Jang-Hee Yoo

Abstract—To increase reliability of face recognition system, the system must be able to distinguish real face from a copy of face such as a photograph. In this paper, we propose a fast and memory efficient method of live face detection for embedded face recognition system, based on the analysis of the movement of the eyes. We detect eyes in sequential input images and calculate variation of each eye region to determine whether the input face is a real face or not. Experimental results show that the proposed approach is competitive and promising for live face detection.

Keywords—Liveness Detection, Eye detection, SQI.

I. INTRODUCTION

ALTHOUGH the recognition performance of biometric system is nowadays quite satisfactory for most applications, much work is still necessary to allow convenient, secure and privacy-friendly systems to be designed. One of the negative implications of increased technological advancement is the ease with which, one can spoof into a biometric identification system. The increase of attack using fake biometric reduces reliability and security of biometric system. Because general biometric algorithms can not be able to differentiate 'live' biometric from 'not live' biometric, the research on liveness detection is highly desirable, yet rather unexplored anti-spoofing measure in biometric identity authentication [1-2].

In face recognition, the usual attack methods may be classified into several categories. The idea of classifying is based on what verification proof is provide to face verification system, such as a stolen photo, stolen face photos, recorded video, 3D face models with the abilities of blinking and lip moving, 3D face models with various expressions and so on [5]. To resist these attack methods, a successful live face detection system should have one or more anti-imposture abilities to expose them. The vein map of the faces using ultra-violet cameras is a most secure method of identifying a live individual, but it needs special expensive devices.

In this paper, we proposed a method of live face detection to resist the attack using a photograph. Our algorithm is based on analysis of movement of facial components, especially eyes, in sequential images. Generally in sequential face images there are very little variations in shape of face and facial components. But eyes have much larger variation in shape because we always blink and move the pupils unconsciously. So we detect eyes in sequential face images and compare the shape of each eye region to decide whether the input face image is a real face or a photograph.

Liveness detection has been a very active research topic in fingerprint recognition and iris recognition communities in recent years. But in face analysis some approaches have been recently presented to deal with this problem [9]. In [3] a depth map is constructed by recovering 3D structure from motion. The depth information can be used to distinguish whether an input face is a live face or a photograph because this depth map is constant in case of a photograph, even if in motion, whereas a live face yields varying depth values. Aggarwal [4] used both optic flow method and feature-based method for estimation of the structure of the image sequence. For optic flow method, he segmented optic flow map and then grouped pixels corresponding to separate objects. Once the flow of each pixel is computed, the 3-D coordinates of surface points can be evaluated. Jain et al. [5] uses the analysis of frequency spectrum of a live face. They define two descriptors to measure the high frequency proportion and the temporal variance of all frequencies. Their method relies on both the lack of quality of a photograph and the change of pose in a live face. However, the above method will be defeated if a very clear and big size photo is used and there is no pose, expression change of user.

In this paper, we don't use any excessive and costly camera (e.g. several cameras including stereo, infrared cameras, etc) or interact with the client demanding real-time responses (e.g. talk, act, etc.) for detecting live face. We propose the methods of live face detection by using analysis of face image sequence captured by one camera as shown in Fig. 1.

Hyung-Keun. Jee is with the Electronics and Telecommunications Research Institute, Daejeon, South Korea (phone: +82-42-860-1674; fax: +82-42-860-5022; e-mail: hkjee@ etri.re.kr).
Sung-Uk Jung is with the Electronics and Telecommunications Research Institute, Daejeon, South Korea (phone: +82-42-860-3849; fax: +82-42-860-5022; e-mail: brcastle@ etri.re.kr).
Jang-Hee Yoo is with the Electronics and Telecommunications Research Institute, Daejeon, South Korea (phone: +82-42-860-1324; fax: +82-42-860-5022; e-mail: jhy@ etri.re.kr).

Fig. 1 Example face image sequence

World Academy of Science, Engineering and Technology
International Journal of Computer and Information Engineering
Vol:2, No:6, 2008

The organization of the paper is as follows. In section 3 our proposed live face detection algorithm is presented. And experimental results are shown in Section 3. Finally, we present our conclusion in Section 4.

II. PROPOSED LIVENESS DETECTION SYSTEM

Fig. 2 shows the overall structure of the proposed liveness detection system. For live face detection we use 5 sequential face images.

```
┌─────────────────────────────┐
│        Input Image          │
└─────────────────────────────┘
              │
              ▼
┌─────────────────────────────┐
│       Eye Detection         │
└─────────────────────────────┘
              │
              ▼
┌─────────────────────────────┐
│  Face Region Normalization  │
└─────────────────────────────┘
              │
              ▼
┌─────────────────────────────┐
│   Eye region Binarization   │
└─────────────────────────────┘
              │
              ▼
┌─────────────────────────────┐
│     Calculate Variation     │
└─────────────────────────────┘
              │
   Yes        ▼        No
        ◇ Var < Thr ◇
  │                       │
  ▼                       ▼
┌──────────┐      ┌──────────┐
│ Live Face│      │ Fake Face│
└──────────┘      └──────────┘
```
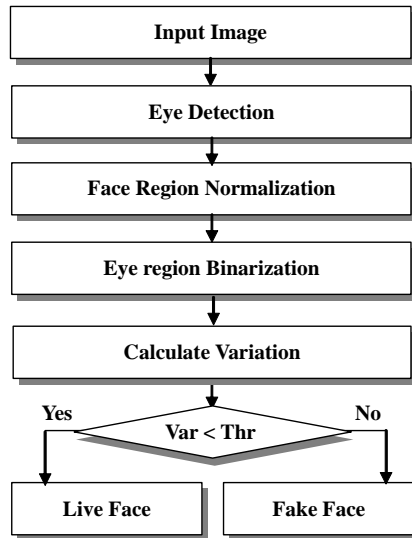
Fig. 2 The flowchart of live face detection

First we detect center point of both eyes in the input face image. Using detected both eyes, we normalize face region and extract eye regions. After binarizing extracted eye regions, we compare each binarized eye regions and calculate variation. If the result is bigger than threshold, the input image is recognized as live face, if not, it is discriminated to the photograph.

A. Eye Detection

Accurate eye detection is very important because we use variation of eye shape to detect live face. This section shows the process to find the eye in the input face image. When the face image is considered as a 3D curve, the intensity of the eye region is lower than the rest of face region. To find the eye candidates using this characteristic, first, we perform Gaussian filtering to the face image, so that the smoothened 3D curve is obtained. In the curve, we extract all the local minimums using the method of the gradient descent [6]. Gradient descent algorithm is an optimization algorithm that approaches a local minimum of a function by taking steps proportional to the negative of the gradient of the function at the current point. If instead one takes steps proportional to the gradient, one approaches a local maximum of that function

To reduce the invalid eye candidates, we used the eye classifier, which is trained by Viloa's AdaBoost training methods [7]. AdaBoost is a simple learning algorithm that selects a small set of weak classifiers from the large number of potential features.

To use above strong classifier, first, we extract the 16×16 sub-windows around each eye candidates during reducing the size of face image. The reason why we use image pyramid is to be robust to different eye size. Then, we insert the sub-windows to the strong classifier. Eyes can be determined by the eye candidates which has the maximum value.

B. Face Region Normalization

Because the input face can vary in size and orientation, we normalize face region about a size and rotation by using center points of both eyes. First we rotate the image with angle of both eyes. After rotating, we extract face region. When distance of both eyes is d, we extract face candidate as much as 1/2d to outside of eyes, 1/2d to upside of eyes, and 3/2d to downside of eyes. Then we normalize this face region to the size of 72x72.

Then we apply SQI to the face region to decrease the effect of illumination. Self Quotient Image (SQI), Q is described in Eq. (1). SQI can be thought as a sort of high pass filter [8].

$$Q = \frac{I}{\hat{I}} = \frac{I}{F * I} \tag{1}$$

where $\hat{I}$ is the low frequency image of the original image, F is the Gaussian kernel.

$$G = \frac{1}{2\pi\sigma^2} e^{-\frac{(i^2 + j^2)}{2\sigma}} \tag{2}$$

$$\tau = Mean(I_\Omega) \tag{3}$$

$$W(i, j) = \begin{cases} 0 & I(i, j) < \tau \\ 1 & otherwise \end{cases} \tag{4}$$

$$F(i, j) = W(i, j)G(i, j) \tag{5}$$

where Ω is the kernel size, I(i, j) is the intensity, F(i, j) is the Gaussian kernel, and W(i, j) is the weight in (i, j).

The weight of filter, which applies to each matrix component independently, has 0 when the intensity of pixels is lower than the mean of filtering region (τ), otherwise 1. Gaussian kernel (F) is obtained by the multiplication between weight (W) and Gaussian filter (G). Fig. 3 shows the original images and the illumination rectified images using SQI. The face images in Fig. 3(a) have a lighting variation. In Fig. 3(b) show the lighting variation is unchanged, and the chrematistics of face is remaining.

(a)

World Academy of Science, Engineering and Technology
International Journal of Computer and Information Engineering
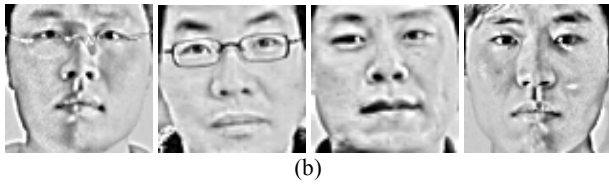Vol:2, No:6, 2008

(b)

Fig. 3 (a) Input gray image (b) Self Quotient Image

C. Eye Region Binarization

After Normalizing face region, eye regions are extracted as 10x20 size based on the center of eyes. Then eye regions are binarized in order to have the pixel value of 0 and 1 by using a threshold. The threshold is adaptively obtained from the mean pixel value of each eye region. Fig. 4 shows the example of binarized eye regions extracted from 5 sequential face images. As shown in Fig. 4 eye regions of fake face change very little, but eye regions of real face have a much larger variation in shape because of blink or movement of pupil.
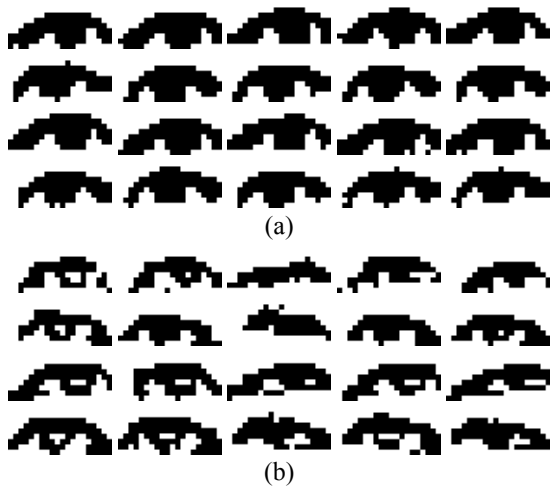

(a)

(b)

Fig. 4 Example of binarized eye regions of fake face (a)
and live face (b)

D. Liveness Score Calculation

Hamming distance method is used to calculate liveness score of each eye region. If two ordered lists of pixels are compared, the Hamming distance is the number of pixels that do not have same value. We compare 5 left eyes each other using hamming distance and 5 right eyes in the same way. The number of pixels which differs between two eye regions becomes liveness score. After adding 10 liveness scores of left eyes and 10 liveness scores of right eyes, we take an average of scores. If the average liveness score is bigger than threshold, we recognize the input image as live face and in the case of opposite it is discriminated to the photograph.

III. EXPERIMENTAL RESULTS

A. Database

Unlike face detection and recognition, there is no public database for live face detection. So we print out face images of 10 persons using photoprinter to make fake faces. The size of photographs is 120x100mm. As shown in Fig. 5, for an experiment of fake faces, we hold a photograph before camera with moving or inclining, and for live faces, 10 persons sit down in front of camera with no facial expression change for the several second.

(a)   (b)

Fig. 5 Example of fake face (a) and live face (b)

Then we obtain 5 frames from each fake and live face and repeat 10 times using the method shown above.
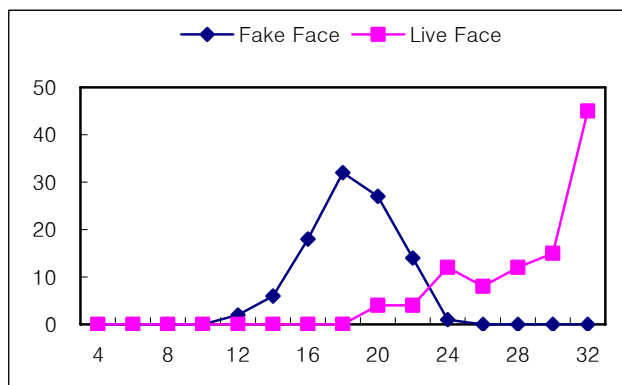
B. Experimental Results

A total of 100 live faces and 100 fake faces were analyzed by the live face detection system described in the previous section. Table I shows the results of liveness score comparing eye regions by using Hamming distance. As shown in Table I, mean score of live face is 30, whereas most of the fake face achieved a score of 17. So we can find that the score of live face is clearly larger than that of fake faces.
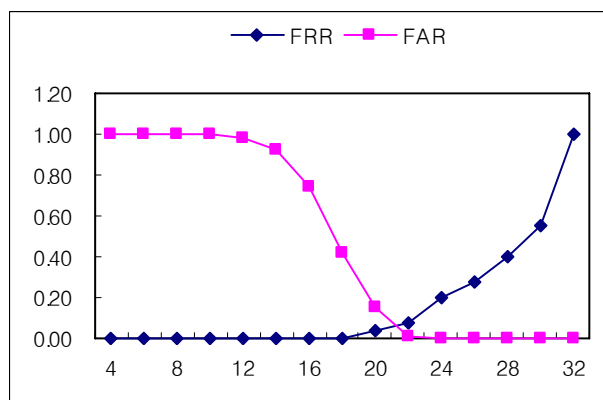
TABLE I
HAMMING DISTANCE OF EYE REGIONS

	Hamming distance		
	Mean	Min	Max
Live face	30	18	47
Fake face	17	10	22

According to the experimental result, when the threshold was set up as 21, we achieved best performance that the FAR is 0.01 (1/100) and FRR is 8 (8/100). Here, the FAR means the error rate of accepting the counterfeit face as the live one. And the FRR does the error rate of rejecting the live face as the counterfeit one.

World Academy of Science, Engineering and Technology
International Journal of Computer and Information Engineering
Vol:2, No:6, 2008

(a)

(b)

Fig. 6 (a) Distribution of liveness scores (b) ROC curve

IV. CONCLUSION AND FUTURE WORK

In this paper, a new method of live face detection to resist the attack using a photograph for embedded face recognition system is presented. Our proposed algorithm is based on analysis of movement of eyes because it has much larger variation in shape among facial components. We believe the results obtained are very promising: the method achieved a reasonable EER (6.5%), proved to be very efficient (on ARM9 400Mhz, the average processing and matching time is less than 400ms). The live face detection system doesn't demand any kind of action (e.g. the change of pose or expression, etc.) to the client. And the proposed approach has also the advantage of being software-based and no additional hardware is required, and hence is economical. As to future research, we intend to investigate a live face detection system which can resist the attack using a video with movement.

REFERENCES

[1] N. K. Ratha, J. H. Connell, and R. M. Bolle, "Enhancing security and privacy in biometrics-based authentication systems," IBM Systems Journal, vol. 40, no. 2, pp. 614-634, 1995.
[2] Stephanie A. C. Schuckers, "Spoofing and anti-spoofing measures," Information Security Technical Report, Vol. 7, no. 4, pp. 56-62, 2002.
[3] T. Choudhury, B. Clarkson, T. Jebara, and A. Pentland, "Multimodal person recognition using unconstrained audio and video," International Conference on AVBPA, pp. 22-28, 1999.
[4] J. K. Aggarwal, N. Nandhakumar, "On the Computation of Motion from Sequences of Images – A Review," Proc. IEEE, vol. 76, pp. 917-935, 1998.
[5] J. Li, Y. Wang, T. Tan, and A. K. Jain, "Live face detection based on the analysis of fourier spectra," In Biometric Technology for Human Identification, SPIE vol. 5404, pp. 296-303, 2004.
[6] R. O. Duda, P. E. Hart, D. G. Stork, "Pattern Classification," 2nd eds, A Wiley-Interscience Publication, 2001.
[7] P. Viola and M. Jones, "Rapid Object Detection using a Boosted Cascade of Simple Features," In Proceedings of IEEE International Conference on Computer Vision and Pattern Recognition, pp. 511-518, 2001.
[8] H. Wang, S. Z. Li, Y. Wang, "Face Recognition under Varying Lighting Condition Using Self Quotient Image," In Proceedings of International Conference on Automatic Face and Gesture Recognition, pp. 819-824, 2004.
[9] W. Zhao, R. Chellappa A. Rosenfeld, P. J. Phillips, "Face Recognition: A Literature Survey," Technical Reports of Computer Vision Laboratory of University of Maryland, 2000.