Live Face Detection

Based on the Analysis of Fourier Spectra

Jiangwei Li [1], Yunhong Wang [1], Tieniu Tan [1], A.K.Jain[2]

1. National Laboratory of Pattern Recognition, Institute of Automation,
Chinese Academy of Sciences, Beijing, 100080 P.R.China;
2. Department of Computer Science and Engineering, Michigan State University,
East Lansing MI 48824 USA

ABSTRACT

Biometrics is a rapidly developing technology that is to identify a person based on his or her physiological or behavioral characteristics. To ensure the correction of authentication, the biometric system must be able to detect and reject the use of a copy of a biometric instead of the live biometric. This function is usually termed “liveness detection”. This paper describes a new method for live face detection. Using structure and movement information of live face, an effective live face detection algorithm is presented. Compared to existing approaches, which concentrate on the measurement of 3D depth information, this method is based on the analysis of Fourier spectra of a single face image or face image sequences. Experimental results show that the proposed method has an encouraging performance.

KEYWORDS: Biometrics Technology, Live Face Detection, Fourier Spectra

1. INTRODUCTION

The recent advances of information technology and the increasing requirement for security have resulted in a rapid development of intelligent personal identification based on biometrics [1,2,3,4]. Biometrics has the capability to accurately distinguish between an authorized person and an imposter. To protect the authentication process, biometric system must be able to reject the use of a copy of a biometric instead of the live biometric [2,3,5,6,7]. This functionality is termed “liveness detection” [6].

In recent years, due to its spacious and successful application, face recognition has received significant attention [8]. More and more researchers in diverse fields, e.g., biometrics, pattern recognition, and computer vision communities, give their attention to face recognition. Face recognition is one of the most active research topics due to its potential applications in access control, automated crowd surveillance, law enforcement, information safety, multimedia communication, human-machine interface, etc. Compared to other biometric authentication technologies, face recognition has an obvious advantage that it does not need much cooperation from users. Just for the broad application foreground of face recognition, how to recognize fake faces is important. This is the problem that live face detection intends to address.

It may be difficult to depict all kinds of face attack methods, because there is no exact and overall description about these methods. Nevertheless, the usual attack methods may be classified into several categories. The idea of classifying is based on what verification proof is provided 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. 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 need special expensive devices. Corresponding to various attack methods, there are various anti-imposture methods.

In this paper, to resist the main fake approach, i.e., using a photo to spoof the face recognition system, a new technique based on the analysis of 2-D Fourier spectra is proposed. Compared to previous methods, our new solution is easier to realize. The algorithm is based on two principles: first, since the size of photo is smaller than that of live face and the photo is flat, high frequency components of photo images must be less than those of real face images. Secondly, even if a photo is held before a camera and is in motion, since the expressions and poses of the face contained in the photo are invariant, the standard deviation of frequency components in sequence must be very small. We use the proposed technique to detect live faces, and experiments are provided to demonstrate its performance.

This paper is organized as follows. In next section, we give a brief review of some related work. Section 3 starts with analyzing 2D spectra of input face images and illustrates our live face detection algorithms. Section 4 presents and discusses experimental results. In Section 5, we conclude the paper.

2. RALATED WORK

Using face patterns as an approach to personal identification and verification can go back to several centuries ago [9], but most of existing work focuses on face detection and recognition. Efforts on live face detection are still very limited, though live face detection is highly desirable. Some papers [3,10] mentioned this topic, however no scheme is designed for live face detection specifically. Here, we will have a brief look at these related works.

Unlike our method, the existing techniques mainly concentrate on the measurement of 3D depth information. The depth information can be used to distinguish whether an input face is form a liveness or a photograph, namely, it is very difficult to fool the system with the ability of estimating head depth information. Previous methods to solve the problem of structure estimation from motion can be roughly divided into two major categories: discrete approach and differential approach [10,11]. In the discrete approach, the depths of a set of features, e.g., points, corners, lines, etc, are computed. Azarbayejani [12] used points where the image intensity has a large Hessian as features. For each frame, a set of features satisfying the Hessian criterion, e.g., eyes, pupils, nostrils, are applied for tracking. EKF (Extended Kalman Filter) is employed to convert the 2-D feature position measurements into 3-D estimates of the position and orientation of the head. It should be noted that if only a camera is available, the structure parameter is represented with an unknown scalar. Choudhary [13] estimated the structure from motion to yield depth estimates for each of the features. He thought the depth value varied greatly for actual faces, while a photo yielded the same depth value. The differential approach is based on computing some field (optical, image and normal) to estimate the depth information. Aggarwal [11] used both optic flow method and feature-based method for estimation of the structure of the image sequences. 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.

However, these algorithms are not for live face detection specifically, but for the computation of structure of objects from a sequence of images, which is combined with the computation of motion [11]. The combined tasks are useful in prediction and segmentation. Moreover, both methods have many disadvantages. For discrete approach, it assumes that the objects undergo rigid motion. Obviously, it is not adaptive for the motion of live face. In addition, it is difficult to establish and maintain correspondence between the features. For differential approach, continuous image acquisition is needed and spatio-temporal gradients are noise sensitive. Such live face detection algorithms are highly computational, too. In this paper, we will propose a low computational cost algorithm to discriminate the live face in frequency domain.

3. PROPOSED ALGORITHM

To be easily forged and peculated is a crucial weakness of traditional personal identification methods. In the same way, biometric characteristics are possible to be forged and peculated. Nowadays, the main attack method is to illegally acquire and use images, video and audio of biometric characteristics of others. For example, utilizing a fingerprint image of the administrator, an imposter may access to a specified system using fingerprints as password. However, biometric characteristics themselves possess some physiological traits which can be explored to effectively prevent such tricks. So, in practical applications, biometrics based identification systems should exactly discover this kind of tricks, namely, correctly detect whether the input data is from a live subject.

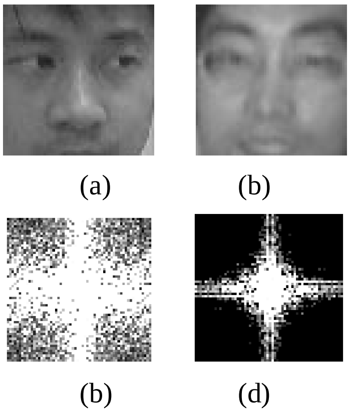
Usually, natural media in which fake faces exist primarily include paper, screen of video device, photo and so on. The structure of these media is greatly different from that of live face. All these media are 2-D planar structure, whereas live face is 3-D structure. According to the Lambertian model [14], the face image can be described as

$$I(x, y) = \rho(x, y) n(x, y)^T s \tag{1}$$

where ρ is the albedo (surface texture) of face, $n(x, y)^T$ is the surface normal (3D shape) of the object (the same for all objects of the class), and **s** is the point source, which can vary arbitrarily. Due to 2-D planar structure of photograph, $n(x, y)^T$ is a constant. So, under the same illumination, images from a liveness are determined by the albedo and the surface normal, whereas those from a fake are determined by the albedo only. We can draw a conclusion from equation (1) that the intensity contrast of live face image is more obvious than that of fake image. Such differences lead to their greatly different reflectivity of light, which is reflected in frequency distribution of an image. Additionally, the size of fake image is usually smaller than that of live face. If they are held before the camera, many details contained in the face captured by the camera will lose. All these bring about great differences between live face image and fake face image, which can be detected by analyzing their 2D Fourier spectra. As shown in Figure 1, compared with the Fourier spectra of a live face image, that of a fake face image has much less high frequency components caused by those factors mentioned above. Hence, to analyze 2D Fourier spectra of the input image is an effective way to live face detection. We calculate the ratio of the energy of high frequency components to that of all frequency components as the corresponding high frequency descriptor (HFD) defined in Equation 2. Obviously, high frequency descriptor of the live face should be more than a reasonable threshold T_{fd}. In this paper, the high frequency components of an image are those whose frequencies are greater than two third of the highest radius frequency of the image and whose magnitudes are also greater than a threshold T_f (generally, the magnitude of high frequency components caused by the forgery process is smaller than that of original image.).

$$HFD=\frac{\iint_{\Omega=\{(u,v)\,|\,\sqrt{u^2+v^2}>\frac{2}{3}f_{\max}\ and\,|F(u,v)|\,>T_f\}}|F(u,v)|\,dudv}{\iint|F(u,v)|\,dudv-F(0,0)}\ \mathrm{x}\,1000 \qquad (2)$$

where *F(u, v)* is Fourier transform of an face image, f_{max} is the highest radius frequency of *F(u, v)*, T_f is a predefined threshold. The denominator means the total energy in frequency domain, namely, the sum of Fourier coefficients relative to direct coefficient. The form can reduce the effect of illumination.


(a) (b)

(b) (d)

Figure 1. Difference between live face and fake face in frequency domain: (a) A live face image; (b) A fake face image; (c) 2D Fourier spectra of (a); (d) 2D Fourier spectra of (b). High Frequency descriptor corresponding to (a) and (b) are 1.6558 and 0 respectively.

However, the above method will be defeated if a very clear and big size photo is used to fool the system. To solve this problem, as similar as previous related works, motion images can be exploited for the live face detection. From video sequences, we can obtain many useful temporal and spatio information. Consider a human who is making a withdrawal before an ATM with a camera, a set of face images will be captured. During this process, the expression and the pose of the liveness are in varying. We assume illumination is unchanged since the whole process takes a short time. But the fake does not have such dynamic characteristic. This is because the expression and the pose of the fake are invariant. Even if the photo is in moving, due to its 2-D planar structure and static characteristic, as long as the face is located very well, the appearances of images are nearly unchanged. So, monitoring temporal changes of facial appearance over time, where facial appearance is represented by an energy value defined in frequency domain, is an effective approach to live face detection. Here, we propose a three-step algorithm to solve this problem. 1) A subset is constructed by extracting image from an input image sequence every four images. 2) For each image in such subset, an energy value *t* defined in Equation (3) is computed. 3) The standard deviation of the resulting flag values, called frequency dynamics descriptor (FDD), is calculated to represent temporal changes of the face. The following is the definition of frequency dynamics descriptor.

$$FDD = \left(\frac{1}{n} \sum_{j=1}^{n} \left(t_j - t_m\right)^2 \right)^{1/2}$$

$$t = \iint | F(u, v) | \, du dv \tag{3}$$

where t_j is the energy value of the jth image, t_m is the mean of energy values, n is the total number of energy values. The energy value t sums up all frequency components and this term presents the global facial gray distribution. Moreover, the variation of the energy value is mainly caused by pose, expression and illumination. Assumed invariant illumination, to photograph image sequences, such variation should to be zero. This algorithm can thus discover such fraudulence as moving a photograph in front of a camera.

Figure 2 shows the energy value curves from four different face image sequences and the corresponding frequency dynamics descriptors. Owing to lack of dynamics, frequency dynamics descriptor of a fake face should be zero and the corresponding energy value curve should be a straight line.

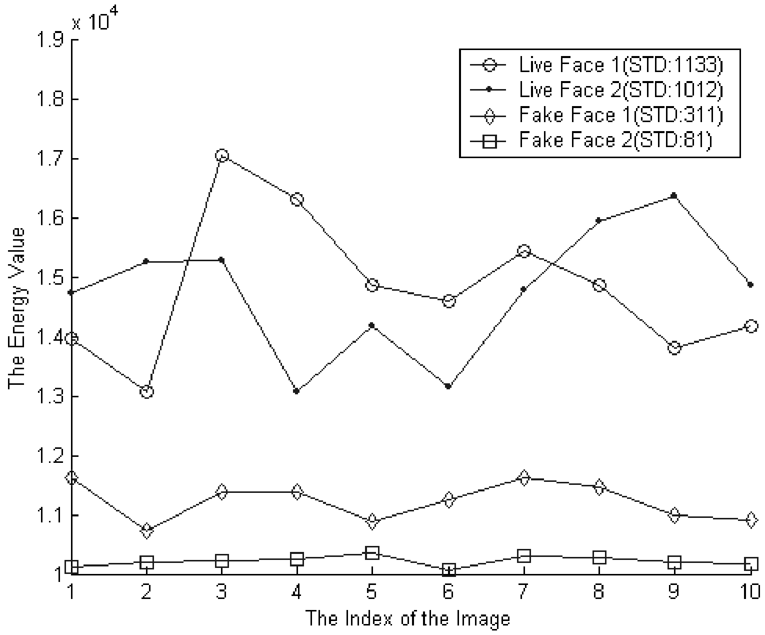


Figure 2. Comparison of four energy value curves between live face and fake face

However, from Figure 2, we can see that the results are not the same as what we expected. The reason is there possibly exists alignment error. However, compared with normal changes of the live face, such alignment error is insignificant. Figure 2 also indicates that the proposed algorithm can well capture the differences between a live face image sequence and a fake face image sequence. Because changes of the energy value of the fake between two successive images or frames are very small (usually, the intervals between two frames is 40ms or less), we compute frequency dynamics descriptor based on a subset of the input images to reduce the computational complexity, instead of the original image sequence.

Based on above analysis on frequency domain, an integrated algorithm is proposed for live face detection. Figure 3 is the flow chart of the proposed algorithm. Considering 3-D structure, varying expressions and poses of live face, the integrated algorithm utilizes both the spatio and the temporal characteristics. When analyzing the high frequency descriptor, we only utilize three images in order to ensure adequate robustness as well as reducing computational cost of the whole algorithm.

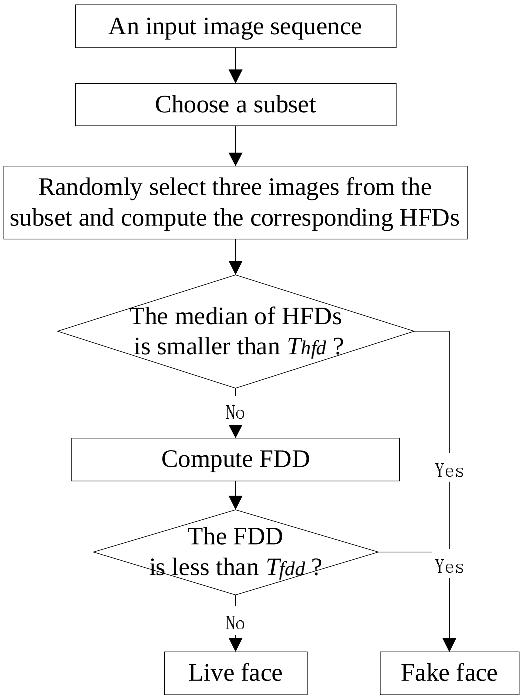
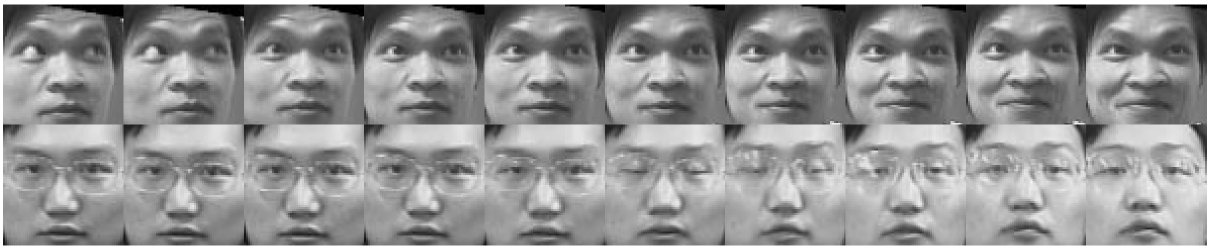

An input image sequence
Choose a subset
Randomly select three images from the subset and compute the corresponding HFDs
The median of HFDs is smaller than T_{hfd} ?
No
Compute FDD
Yes
The FDD is less than T_{fdd} ?
Yes
No
Live face
Fake face

Figure 3. The flow chart of live face detection

4. EXPERIMENAL RESULTS

Unlike face detection and recognition, there is no public database for live face detection. So we construct a face image database to evaluate the performance of the integrated algorithm. This face image database is obtained using a Logitech QuickCam Pro 4000 camera. For fake faces, there are 20 sequences corresponding to four different objects. Each sequence contains 10 frames. Each frame in database is extracted from an original sequence every four images. The media of three objects are photographs and the other is a print paper. The sizes of photographs are 48x33mm, 76x55mm and 124x84mm, respectively. The resolution level of the printed images is 600dpi. Holding a fake face before the camera, we move it according to some styles: moving quickly, moving slowly or inclining. For live faces, there are 20 live face sequences corresponding to four different persons. The subjects sit down before the camera and move as they wish. The length of each live face sequence is 10 frames. Some examples of live and fake face sequences are shown in Figure 4. The first and second rows are from live face. The third and fourth rows are from two photographs which sizes are 48x33 and 76x55mm, respectively. The last row is from a print image. These faces are cut manually. Table 1 lists statistics of the frequency dynamics descriptors and high frequency descriptors of these image sequences.



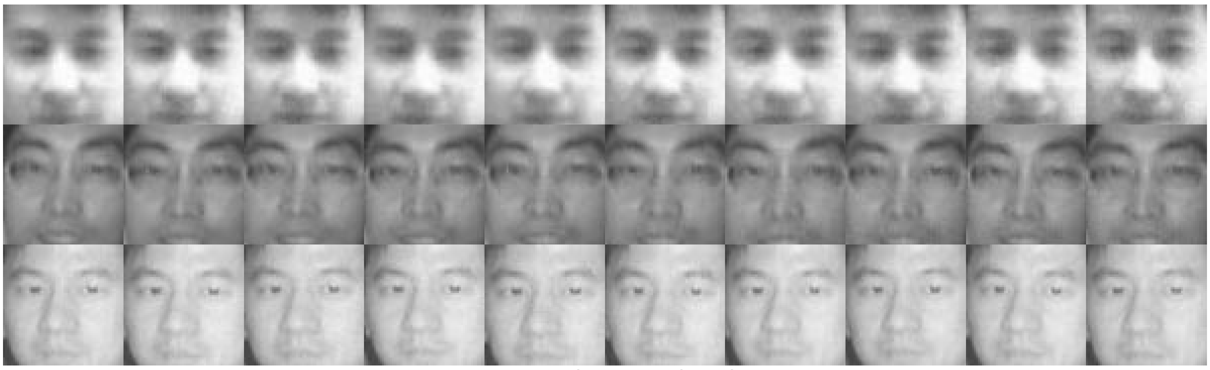


Figure 4. Some examples of live and fake face image sequences

There are two thresholds, T_{hfd} and $T_{fdd,}$ to be confirmed. We choose $T_{hfd} = 0.4$ and $T_{fdd} = 500$. In our experiments, our algorithm can correctly detect all live and fake face image sequences. The performance of frequency dynamic descriptor is superior to that of high frequency descriptor. We notice that for small size fake images, the values of the high frequency descriptor are far below T_{hfd}, but for big size fake images, i.e., 124x84mm, the values of them are nearly or over T_{hfd}. This is because big size photo has more facial details to fool the system. The drawback can be overcome by the frequency dynamic descriptor. As long as faces are located very well, frequency dynamic descriptors can easily determine the liveness of the face. Table 1 shows that the frequency dynamic descriptor of live face image sequences are clearly larger than T_{fdd}. The reasons why the frequency dynamic descriptor of fake faces is not zero are as follows: 1) There are some location errors. 2) Illumination is not a constant as we assumed. We also learn from the table that bigger size faces have lower frequency dynamic descriptors for the reason that they are easier to be located precisely than small size faces. Though the frequency dynamic descriptors are more robust, the high frequency descriptor is very helpful to prevent the small size fake faces, which is used mostly. Moreover, the computation cost of this descriptor is very low. The frequency dynamic descriptor needs good location of faces, which is a difficulty per se.

Image Sequence		Frequency Dynamics descriptor			High Frequency descriptor		
		Mean	Min	Max	Mean	Min	Max
Live face	200 images	960	718	1490	0.7197	0.4011	2.0544
Fake face	40 images (48x33mm)	286	233	376	0	0	0
	50 images (76x55mm)	260	186	364	0.0913	0	0.1376
	90 images (124x84mm)	175	91	282	0.3535	0	0.5514
	20 images (600dpi)	249	237	260	0.2803	0	0.3917

Table 1. Experimental results of live face detection

5. CONCLUSION

In this paper, we have described a new integrated method for live face detection. Live face detection is implemented by analyzing high frequency descriptor and frequency dynamics descriptor of face images, which exhibit physiological characteristics of the live face. High frequency descriptor is very effective to prevent the spoof of small size fake images. Frequency dynamics descriptor is more successful to discover such fraudulence as moving a large size fake image in front of a camera. Experimental results have shown that the proposed method has an encouraging performance. Compared to the previous works, which seek to acquire 3-D depth information of the head, our algorithm has advantages on computation and easiness. In future, we will verify and improve this algorithm based on an enlarged database. In addition, more robust algorithms are needed so as to decrease the effect of illumination changing and execute full automatically.

Acknowledgments

This work is funded by research grants from the National Natural Science Foundation of China (NSFC No. 60121302 and No. 60332010).

REFERENCES

[1] P.Wilson, "Biometrics: Here's Looking At You", <*http://www.insight.co.uk/presscoverage.htm*>.

[2] V.Matyas, Z.Riha, "Biometric Authentication Security And Usability ", <*http://www.fi.muni.cz/usr/matyas/*>.

[3] A.K.Jain and A.Ross, "Biometrics", <*http://www.win.tue.nl/~henkvt/Biometrics*>.

[4] W.T.Freeman, "Book Review Biometrics", <*http://www.merl.com/reports/docs* >.

[5] NK.Ratha, JH.Connell, RM.Bolle, "Enhancing security and privacy in biometrics", <*http://www.research.ibm.com/journal/sj/403/ratha.html*>.

[6] J.Vacca, "Privacy Enhanced Biometrics-Based Authentication", <*http://www.informit.com*>.

[7] B.Cukic, "Introduction to Biometrics", <*http://www.biometrics.org/html/introduction.html*>.

[8] X.Lu, "Image Analysis for Face Recognition", <*http://www.cse.msu.edu/~lvxiaogu/publications/*>.

[9] Y.Fang, "Face Detection and Recognition", PhD thesis, *Institute of Automation, Chinese Academic of Science*, 2003.

[10] 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.

[11] J.K.Aggarwal, N.Nandhakumar, "On the Computation of Motion from Sequences of Images – A Review", *Proc. IEEE*, 1988, **76**: 917– 935.

[12] A.Azarbayejani, T. Starner, B. Horowitz, A. Pentland, "Visually controlled graphics", *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 1993, **15**(6): 602 – 605.

[13] T.Choudhury, B.Clarkson, T.Jebara, A.Pentland, "Multimodal Person Recognition using Unconstrained Audio and Video", *In Proceedings of the 2nd International Conference on Audio-Visual Biometric Person Authentication*, 1998.

[14] R.Basri, D.W. Jacobs, "Lambertian Reflectance and Linear Subspaces", *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2003, **25**(2): 218 – 233.