Face Anti-spoofing Based on Motion

Ran Wang[1], Jing Xiao[1,3,4](✉), Ruimin Hu[1,2,3], and Xu Wang[1]

[1] School of Computer Science, National Engineering Research Center
for Multimedia Software, Wuhan University, Wuhan 430072, China
{wangran1994, jing}@whu.edu.cn,
hurm1964@gmail.com, wangxu9191@gmail.com
[2] Hubei Key Laboratory of Multimedia and Network Communication
Engineering, Wuhan University, Wuhan 430072, China
[3] Collaborative Innovation Center of Geospatial Technology,
Wuhan 430079, China
[4] Research Institute of Wuhan University in Shenzhen, Shenzhen, China

Abstract. People can have access to biometric system easily by face spoofing attack. Recent researches have proposed many anti-spoofing strategies based on eye blinking, facial expression changes, mouth movements or skin texture. As for the face which has slight trembling, there are few specific methods about it. To solve this problem, we have proposed a method which could discriminate human real face and face print by using parameters of slight face motion and equal-proportion property in projection. In order to validate our method, we also established a new video database containing 720 moving faces. After getting the facial landmarks, aligning image of each frame and calculating the variance as feature value, final data would be sent to the support vector machine (SVM) to verify the reality of faces. From the experiment results, the proposed method shows its high accuracy in different lighting conditions and different face amplitude for anti-spoofing and will have good prospect in engineer.

Keywords: Anti-spoofing · Face · Motion · Projection · Classification

1 Introduction

Biometric system which relies on biometric identifier taken from the inherited traits of user, has developed rapidly in applications recently [1, 2]. With the widespread use of biometric system, many kinds of attack which try to get pass access has arisen.

Unfortunately, it is proved that biometric system can't prevent the face spoofing attack because the biometric system is based on the biometric identities, it doesn't consider the liveness of identity [3]. The most common kind of face spoofing attack is print attack, this attack performed using a face print in front of the collecting device of biometric system to disguise the real identity of imposter [4]. Beside this, face spoofing attack also includes video attack and mask attack [5].

Liveness detection is the typical countermeasure of face spoofing attack. It means that facial physiological changing such as eye blinking, facial expression changes and mouth movements can all be used as the characteristics for the face anti-spoofing methods [6]. In recent research, lots of related papers proposed their methods in many

B. Zeng et al. (Eds.): PCM 2017, LNCS 10736, pp. 202–211, 2018.
https://doi.org/10.1007/978-3-319-77383-4_20

kinds of aspects, but most of the methods were based on the same situation: the detected faces should be in front of the camera without any trembling. In order to adjust to this situation, before the experiments, researchers always chose the appropriate faces which collected from public databases such as NUAA (Fig. 1) [1] and PRINTATTACK [7] or the database developed on their own. In these databases, a client is required to stay still, just blink or roll eyes, smile or open mouth. These experiments could usually give us satisfactory results.

But these proposed methods have the same limitation. On the one hand, consider this situation, if a client doesn't follow the instruction which let him be still or an attacker put a face print in front of the camera with trembling because of the cold weather, it is difficult to collect the accurate facial information, and these methods seem to make the final discrimination hardly because most papers don't consider this situation. If the algorithm of the method discriminate the human as attacker, it will be too arbitrary. In fact, this situation is usual in the normal life; On the other hand, slight trembling of face print can really cause attack because the motion of a print image such as translation and rotation will disguise itself as a three dimensional (3D) face.

In order to solve this problem, in this paper, we proposed a new face anti-spoofing method by using parameters of face slight trembling motion and the differences projection characteristics between plane image and 3D subject. We considered the faces collection process as a projection: for a face print, projection has the property named equal-proportion. It means print face will have special characteristic in projection. When we move or rotate the print face in front of a camera, it corresponds to project the image multiple times. This will make obviously different feature parameters which can be easily classified by the SVM [8].

The development of a new method to solve the problem is not possible without an appropriate database. As we mentioned above, almost all the databases contains abundant print attack images or videos, but few of them have the subjects with face trembling. It will limit our experiment if we choose the public database. To solve this limitation, we developed a new database, which consisted of client dataset and imposter dataset. Each video clip has the face sequences we need. To improve the persuasiveness and accuracy of our database, besides the consideration of different lighting conditions, we invited 60 people in different ages and not only collected the videos with trembling of face, we also collected still face videos like normal print attack database such as PRINTATTACK.

The rest of the paper is organized as follows. Section 2 discusses related works on countermeasures of face spoofing attack. Section 3 gives the details of our proposed method. Our developed database, which consists of client and imposter datasets are introduced in Sect. 4, the experiment of our proposed method in different conditions and the results of them are also given in this section, the results are thoroughly analyzed and also compared with each other. A conclusion is drawn in Sect. 5.

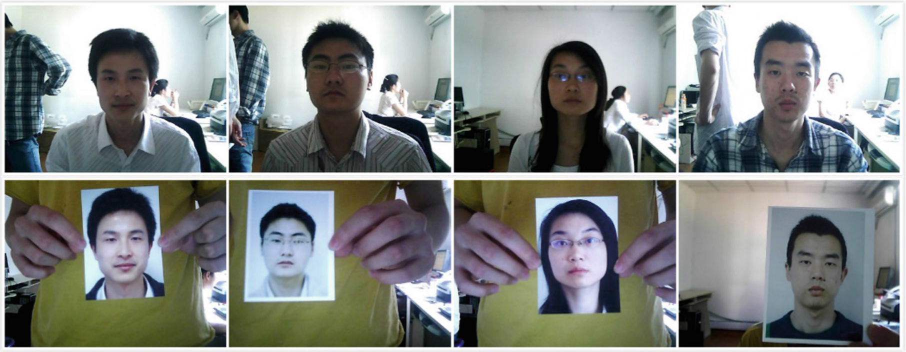

Fig. 1. Example of images captured from real faces (upper row) and from face prints (lower row) in NUAA database

2 Related Work

Face anti-spoofing is a popular research direction in recent years after the biometric system proposed. A lot of related works have been carried out to find effective methods in order to resist face spoofing attack.

As we mentioned above, liveness detection is the typical countermeasure of face spoofing attack [6]. The most basic features of liveness are eye blinking and mouth movements. This is because these facial expressions are very difficult to be imitated [9]. According to this features, for instance, as for eye blinking detection, Pan et al. [10] proposed a face anti-spoofing method based on eye blinking, they collected and detected eye blinking of human in a few seconds and found the differences between real human face and face print [6]. Of course, these physiological signs could be used together, Roberto et al. [11] made the fusion of multiple clues, and analyzed both video and static information to obtain excellent effect of discrimination.

There are still other method based on different theory in face anti-spoofing, Andre et al. [7] thought a public database and a baseline could be really useful to resist face spoofing attack. Jukka et al. [6] used micro-texture analysis method by local binary patterns (LBP) [12] to solve the attack problem. Local ternary pattern (LTP) [13] method in texture analyze of face anti-spoofing also gained wide range of applications. Optical-flow theory was also presented to capture subtle movement and velocity, because the captured optical-flow was obviously different between real human face and face print [6, 14]. For instance, Bao et al. used optical-flow for motion estimation and detected attacks which produced with planar media, the experiments showed a 6% false-alarm against about 14% false-acceptance [15].

It appears that most of the existing methods for spoofing detection not consider the motion differences between the real human face and face print. In fact, the liveness information detection can not only find in the almost still face with little eye blinking motion, when face has some motion, the differences are more obvious. Therefore, we proposed a new kind of discrimination method by using parameters of face slight motion in this paper.

3 Proposed Method

This section provides details of our proposed method. The general overview of our method is summarized in Fig. 2.

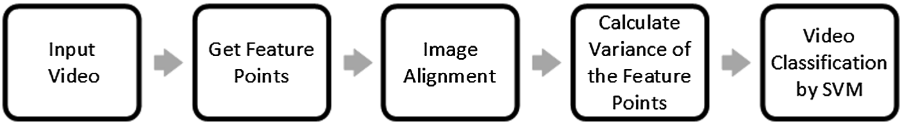
Input Video
Get Feature Points
Image Alignment
Calculate Variance of the Feature Points
Video Classification by SVM

Fig. 2. General overview of our proposed method

As we all know, one of the property of projection is its equal-proportion. It means that for a plane figure, whether the projection direction, if point M is the midpoint of a line segment AB in origin graph, the projection point of the M named M' is also the midpoint of the projection line segment $A'B'$ (Fig. 3).

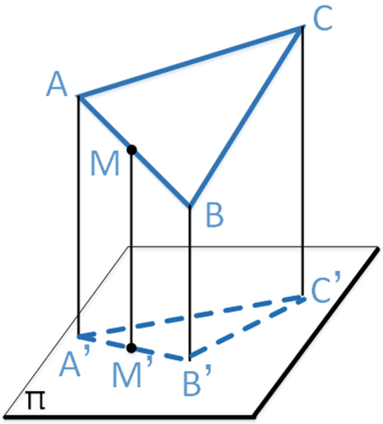

Fig. 3. Equal-proportion of projection, after the projection, it still has the equation $AM : MB = A'M' : M'B'$

In our proposed method, we tried to find obvious evidence to discriminate real faces and face prints by using the characteristic of projection we mentioned above. When we put a face print in front of a camera, it corresponded to the projection from the print. Prints trembling could be seemed as motion. Moved or rotated the prints could be equivalent as changing the direction of the projection. We could also seem it as a still print but it would be projected on a moved or rotated projection surface. Because of the equal-proportion of projection, all points on the print would have the same motion mode.

As for real face, points on the face didn't have the same deep information because the points were not at the same surface. It would cause the disparity of motion mode among all the points. We described the motion mode by homography matrix.

Attention the flowchart in Fig. 2, for the given video, we first got the feature points of the face. The face alignment algorithm in SeetaFaceEngine[1] would help us get some facial landmarks in each frame; then we used three of them to do radiation transformation and calculated homography matrix between the first frame and the others; after that, the left points we had should be aligned with the same points in the first frame by using the homography matrix. The deviation value of the left points should be calculated if they didn't overlap, and we would get the variance from the whole video frame sequences; at last, by using the high dimensional variance, we could get the discriminate result from SVM classification.

3.1 Feature Points Processing

Feature points as the initial value are necessary for the final discrimination of the face. For the proposed method, we chose the SeetaFaceEngine to obtain the feature points in each frame. We could get n facial landmarks by the SeetaFace Alignment modular, such as two eye centers, one nose tip, and two mouth corners. Denoted their coordinates by $p_i(x, y), i = 1, 2, \ldots, n$.

In our proposed method, we made the parameters of face slight motion more intuitive. Consider seven feature points we got, for each frame, used three of them (two eye centers p_1, p_2 and one nose tip p_3) to calculate the homography matrix (H matrix) with the first frame, the SeetaFaceEngine would help us find these points in each frame. Then let the left points ($p_4, \ldots, p_7$) and H matrix do affine transformation. This step would make every frame align with the first frame, and it was easy to know that $p_4, \ldots, p_7$ of the latter frame could hardly overlap the first frame because of the face transformation. The updated four points could be denoted by $p_4^{'}, \ldots, p_7^{'}$.

3.2 Calculate of High Dimensional Variance

In order to judge reliability of the given face in the video, some defined parameters should be calculated for the final classification. At the first stage, we used facial information to connect every follow-up frame with the first one.

As we mentioned at first, real human faces and face prints showed different properties when faces had any slight transformation. For each frame of the video, a real human face could be seen as a still picture while a face print could be seen as a picture-in-picture. This would make the left four points have obviously deviation in the whole video frames. The deviation could be described by the variance of the video frame sequences. To simplify the process of variance calculation, we selected 100 serial frames of the whole video frame sequences randomly. For the extracted 100 frames, $p_m^n(x)$ present the x direction value of the No. m feature point in the No. n frame.

[1] https://github.com/seetaface/SeetaFaceEngine.

Denote the variance by $\sigma_{ij}, i = 4, 5, 6, 7; j = x, y$. Clearly, we will have eight σ data to send to the SVM finally. For example, σ_{4x} can be defined as:

$$\sigma_{4x} = \mathrm{D}(p_4^n(x))|n = 1, 2, \ldots, 100 \tag{1}$$

Where D means the variance.

4 Experiments and Analysis

In this section, we will introduce the experiment details of our proposed method and discuss about the results in order to do some analysis. We begin with the introduction of our developed database. Then we will show the experiment results by some intuitive charts. The results will be given by some comparisons among different conditions of experiment.

4.1 Database

We developed a database which contained the human faces and face prints video clips in three lighting conditions: strong light, moderate light and weak light. According to the amplitude of the face motion, video clips in each condition were divided into two kinds of group. The general overview of the database is given in Fig. 4.

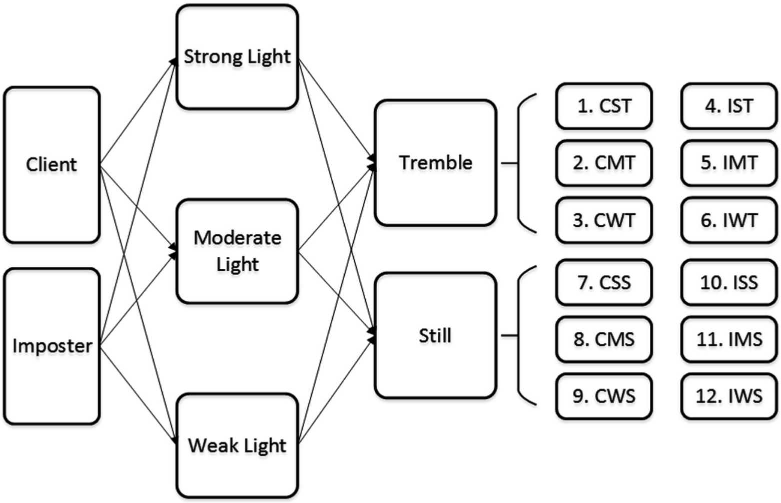

Fig. 4. Overview of our developed database which is classified by twelve labels. The naming convention for labels is using acronyms of the each experiment condition, each label has 60 face video clips

For the real faces (client) dataset, we invited 60 people to participate in the video recording. People were consisted of different ages and different gender. The amount of each group were basically equal. In each video, the clients needed to turn their faces left and right to stimulate trembling, the amplitude was controlled in ±15°. The whole video would be last for about 10 s. We chose the Canon camera to make video and controlled the resolution in 270 × 480 and the filming distance in 50 cm.

For the face prints (imposter) dataset, in order to ensure the effectiveness and reliability of the whole database, we chose all clients in real face dataset, and took their photograph samples using a white curtain as the background. Then put them in front of the camera, did the same motion like the real faces dataset. The face area should take at least 2/3 of the whole area of the photograph. Also kept the video resolution in 270 × 480 and the filming distance in 50 cm.

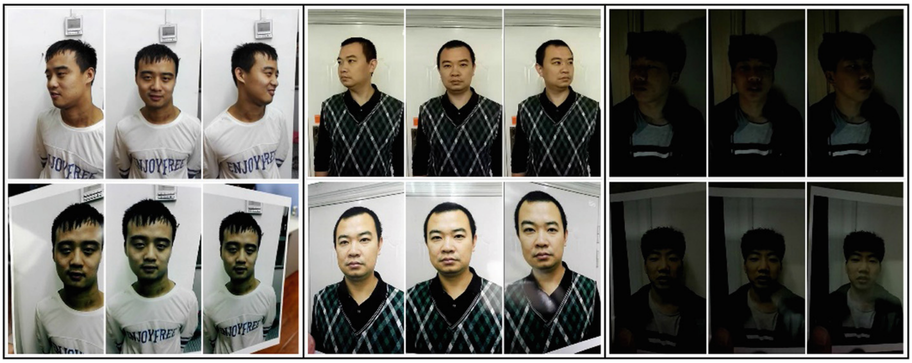

Fig. 5. Example of images captured from real faces (upper row) and from face prints (lower row) in our developed database. The three clients are stay in different lighting conditions, first one (the first three columns) is in strong light, second person (the middle three columns) is in moderate light, and the third person (the last three columns) is in weak light. All of them are still or have slight face trembling.

The capture images of our developed database in different light conditions is given in Fig. 5.

Since the result of the classification had just two possibilities, and there were eight feature vectors (eight σ), the SVM we chose was high dimensional SVM. The kernel type was linear to avoid over fitting.

4.2 Experiment Results and Discussion

From the labels in the Fig. 4, we emphatically considered the effect of our proposed method in different lighting conditions and in different amplitude of the face motion. For the whole database, including the client dataset and imposter dataset, we selected 30 videos from each label as training data (360 videos as training data in all) and the rest of the videos were seemed as test data. Then selected 7 facial landmarks by SeetaFace Alignment modular ($n = 7$).

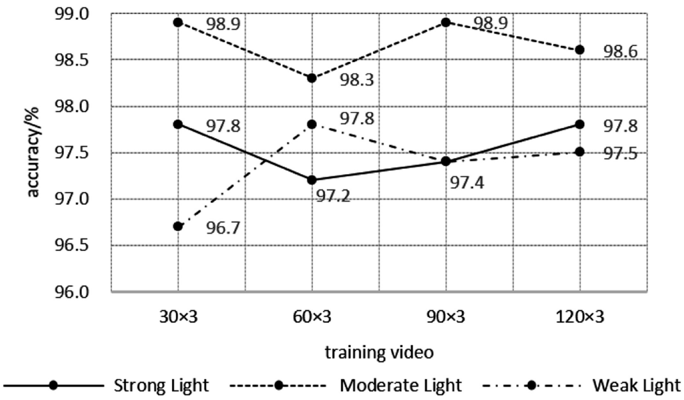

Fig. 6. Comparison of accuracy among different lighting condition. Strong light condition labels (*S*): CST, IST, CSS, ISS. Moderate light condition labels (*M*): CMT, IMT, CMS, IMS. Weak light condition labels (*W*): CWT, IWT, CWS, IWS. 30 × 3 means selecting 30 videos randomly from every light condition label

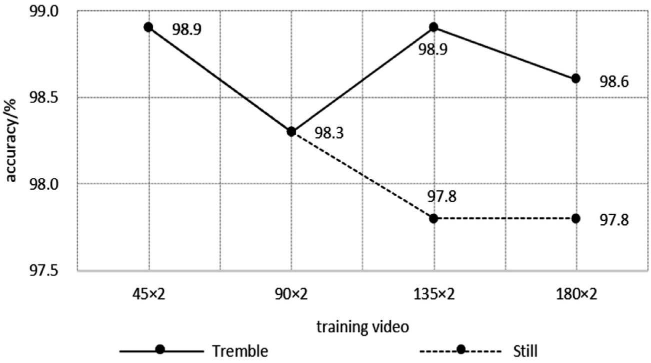

Fig. 7. Comparison of accuracy between different motion amplitude. Face trembling labels (**T): CST, CMT, CWT, IST, IMT, IWT. Moderate light condition labels (**S): CSS, CMS, CWS, ISS, IMS, IWS. 45 × 2 means selecting 45 videos randomly from every face motion amplitude condition label

The comparison of accuracy among different lighting conditions is given by the line chart in Fig. 6 and this criteria between different motion amplitude is also given in Fig. 7.

Since the current anti-spoofing methods always use images rather than videos to test the authenticity of faces, we didn't compare the accuracy of our proposed method with other anti-spoofing methods, our method need the face video clips to calculate the homography matrix.

From the line charts of different experiment conditions, we can get the accuracy percentage of our proposed method. In general, the method results can all get above 96% of accuracy in any conditions. For the light condition, the method shows an excellent performance in moderate light, because whether the light strong or weak, it would influence the face detection and made the facial landmarks getting hardly. From Fig. 6, the detection accuracy in strong light and moderate light decrease first and then increase with the increase of training videos, but it is opposite in the case of weak light. This is because in weak light condition, the collection ability of camera decreased, the detection accuracy had slightly fluctuation. But this fluctuation just remained within the range of 1%. For the face motion amplitude condition, we can see that face in video sequences which has slight trembling motion got good discrimination capacity than the face which nearly still: our method emphatically considered the parameters of face slight motion, and used the homography matrix to calculate the variance value of 100 frames, this would cause less effective of nearly still face since the homography matrix would be calculated inaccurately.

5 Conclusion

In this paper, we found that current face anti-spoofing methods had limitations in face motion situation. So we proposed a face anti-spoofing recognition method that worked on print attack using projection characteristic differences between 3D real face and plane face print. The method was based on a simple but useful optical principle. By SVM classification, the experiment shows efficient recognition result in different conditions.

Furthermore, video clips in various environment of our developed database simulate different situations in reality. The experiments demonstrated that our method can be used in almost extreme illuminating environment or motion amplitude of the face. Another direction to carry forward this work would be to make our method adjust to more situation and to enhance it robustness.

Acknowledgements. This work was partly supported by Hubei Province Technological Innovation Major Project (No. 2016AAA015), the National Nature Science Foundation of China (61502348), the EU FP7 QUICK project under Grant Agreement No. PIRSES-GA-2013-612652, and the science and technology program of Shenzhen (JCYJ20150422150029092).

References

1. Tan, X., Li, Y., Liu, J., Jiang, L.: Face liveness detection from a single image with sparse low rank bilinear discriminative model. In: Daniilidis, K., Maragos, P., Paragios, N. (eds.) ECCV 2010. LNCS, vol. 6316, pp. 504–517. Springer, Heidelberg (2010). https://doi.org/10.1007/978-3-642-15567-3_37
2. Jain, A.K., Flynn, P., Ross, A.A.: Handbook of Biometrics. Springer, USA (2008)
3. Chingovska, I., Anjos, A., Marcel, S.: On the effectiveness of local binary patterns in face anti-spoofing. In: Biometrics Special Interest Group, pp. 1–7. IEEE (2012)
4. Galbally, J., Marcel, S.: Face anti-spoofing based on general image quality assessment. In: International Conference on Pattern Recognition, pp. 1173–1178. IEEE (2014)

5. Erdogmus, N., Marcel, S.: Spoofing in 2D face recognition with 3D masks and anti-spoofing with Kinect. In: IEEE Sixth International Conference on Biometrics: Theory, Applications and Systems, pp. 1–6. IEEE (2013)
6. Maatta, J., Hadid, A., Pietikainen, M.: Face spoofing detection from single images using micro-texture analysis. In: International Joint Conference on Biometrics, pp. 1–7. IEEE Computer Society (2011)
7. Anjos, A., Marcel, S.: Counter-measures to photo attacks in face recognition: a public database and a baseline. In: International Joint Conference on Biometrics, pp. 1–7. IEEE Computer Society (2011)
8. Duda, R.O., Hart, P.E., Stork, D.G.: Pattern Classification, pp. 119–131. Wiley, New York (2001)
9. Pan, G., Sun, L., Wu, Z., et al.: Monocular camera-based face liveness detection by combining eyeblink and scene context. Telecommun. Syst. **47**(3–4), 215–225 (2011)
10. Pan, G., Wu, Z., Sun, L.: Liveness detection for face recognition. In: Recent Advances in Face Recognition. InTech (2008)
11. Tronci, R., Muntoni, D., Fadda, G., et al.: Fusion of multiple clues for photo-attack detection in face recognition systems. In: International Joint Conference on Biometrics, pp. 1–6. IEEE Computer Society (2011)
12. Ojala, T., Pietikäinen, M., Mäenpää, T.: Multiresolution gray-scale and rotation invariant texture classification with local binary patterns. In: European Conference on Computer Vision, pp. 404–420. Springer-Verlag (2000)
13. Parveen, S., Ahmed, S., Mumtazah, S., et al.: Texture analysis using local ternary pattern for face anti-spoofing. Sci. Int. **28**(2), 968–970 (2016)
14. Kollreider, K., Fronthaler, H., Bigun, J.: Non-intrusive liveness detection by face images. Image Vis. Comput. **27**(3), 233–244 (2009)
15. Bao, W., Li, H., Li, N., et al.: A liveness detection method for face recognition based on optical flow field. In: International Conference on Image Analysis and Signal Processing, pp. 233–236. IEEE (2009)