Face Spoofing Detection for Smartphones using a 3D Reconstruction and the Motion Sensors*

Kim Trong Nguyen[1], Cathel Zitzmann[2], Florent Retraint[3], Agnès Delahaies[4],
Frédéric Morain-Nicolier[4] and Hoai Phuong Nguyen[4]

[1] *EPF Graduate School of Engineering, Troyes, France*
[2] *EPF Graduate School of Engineering, Montpellier, France*
[3] *LM2S - ICD - Troyes University of Technology, Troyes, France*
[4] *CREsTIC, University of Reims Champagne-Ardenne, Reims, France*

Keywords: Digital Forensics, Spoofing Attack, 3D Reconstruction, Facial Recognition, Motion Sensors.

Abstract: Face recognition system is proven to be vulnerable to face spoofing attack. Many approaches have been proposed in the literature to resolve this vulnerability. This paper proposes a novel method dedicated to mobile systems. The approach asks users to capture a video by moving the device around their face. Thanks to a 3D reconstruction process, the shape of the object is estimated from the video. By evaluating this 3D shape, we can rapidly eliminate attacks in which a photo of a legitimate face is used. Then, the camera's poses estimated from the 3D reconstruction is used to be compared to the data captured from the device's motion sensors. Experimental results on a real database show the efficiency of the proposed approach.

1 INTRODUCTION

Authentication by facial recognition can be exploited as an additional solution to reinforce the security level of our information systems. However, it is proven that this solution is vulnerable. Facial recognition is easily compromised by face spoofing attacks. Therefore, photos and video widely shared on social networks may become a weapon against their owner's security.

Attackers have many ways to attack a facial recognition system. They can utilize a photo of legitimate user printed on a piece of paper or displayed on an LCD screen and present it in front of the camera in operation. They can also replay a video which filmed the victim previously or just use a 3D mask to mislead the face detection process.

In 3D-mask attack, attackers have to focus on their target and do firstly manage to construct a 3D mask or maybe a sculpture of the target. If the mask is constructed perfectly, there is less chance to detect it. However, the achievement of this type of attack is quite difficult and expensive. In this paper, the proposed method seeks to detect basically the photo and video-replay attacks.

2 RELATED WORKS

Many approaches have been proposed in the literature to deal with face spoofing attacks using different features like texture, liveliness, structure, etc. Textural information, which manages to be different between real-face images and fake ones, can be exploited for face spoofing detection. From a single image, in (Maatta et al., 2011), the authors propose to analyze the texture of facial images using multiscale Local Binary Pattern (LBP). In the same spirit, Kim *et al.* (Kim et al., 2012), also utilized LBP, but in fusion with frequency analyses by using the power spectrum. Other researchers exploited the Local Graph Structure (LGS) (Housam et al., 2014) or its improved versions (ILGS, SLGS) (Abdullah et al., 2014), (Bashier et al., 2014) as texture descriptors to conceptualize their face spoofing detection method. Another method (Nguyen et al., 2016) proposes to exploit the statistic behavior of the distribution of noises local variances to detect face spoofing attacks.

Some approaches manage to distinguish real faces from spoofed faces by seeking proofs of liveness from a sequence of images or from a video capturing the face. Kollreider *et al.* (Kollreider et al., 2007) proposed an approach in which lip movements are ex-

*This work is funded by Troyes Champagne Metropole.

Nguyen, K., Zitzmann, C., Retraint, F., Delahaies, A., Morain-Nicolier, F. and Nguyen, H.
Face Spoofing Detection for Smartphones using a 3D Reconstruction and the Motion Sensors.
DOI: 10.5220/0006602202860291
In *Proceedings of the 4th International Conference on Information Systems Security and Privacy (ICISSP 2018)*, pages 286-291
ISBN: 978-989-758-282-0

ploited for face spoofing detection while the user is asked to speak some numerical digits. Huyng-Kean Jee *et al.* (Hyung-Keun et al., 2006), in their approach, proposed to study uncontrollable movements of eyes regions, such as eye blinking or pupil movement. Following the same idea, Lin Sun *et al.* (Sun et al., 2007) also proposed to exploit eyes movements by modeling and detecting the two principal states of the eyes: opened-state and closed-state.

Some other methods exploit the differences between 2D objects and a 3D face in their structure, their moving features or the depth information that they provide. For instance, Kim *et al.* (Kim et al., 2013) proposed to compare images captured in different focusing. For a 3D object, due to depth information, the difference between images of different focusing will be clearer than the one in the case of 2D objects. The approach permits to identify efficiently spoofing attacks using a 2D display support. Studying the difference in the behavior of optical flows generated by 2D spoofed face and real face have been also envisaged (Bao et al., 2009).

TODO more related works

3 PROPOSED APPROACH

In the last few years, we can remark a constant evolution of mobile technology and of the smartphone market. More and more people use smartphones to ease their daily life as well as their professional activities. Myriad mobile applications require or have access to personal or private information of users. Therefore, they need a high level of security. Authentication by facial recognition is proposed as a solution to reinforce the security of mobile systems. However, the problem of face spoofing is always unavoidable. Actual solutions are quite relevant and optimized to settle this problem, but just in some provided cases study. Thus, an efficient solution dedicated to smartphone system is indispensable.

In the case of a smartphone system, which is mobile, images or videos could be captured under different conditions of lighting, under different orientations and with an uncontrollable background. The quality of acquisition could also be affected by the movement of the camera and the movement relative of the context where situated the acquisition system (e.g. when a user authenticates while he is traveling in a train). In addition, the diversity and the constant evolution of smartphone models, as well as the difficulty in calibrating their cameras, are also among the big barriers for an efficient face spoofing detection solution.

However, the presence of different sensors integrated into a smartphone may be an advantage which allows us to develop a novel dedicated solution to face spoofing detection. Indeed, with the help of the movement sensors and the multitasking ability of smartphones, we can simultaneously capture the device's movement information while filming the user's face by our Android Application. (Notice that all smartphones in our day include at least the gyroscope sensor.) At the end of this phase, the output will include a video of the head and sensors raw data. In the case of legit authentication, information given by movement sensors is a priori coherent with information estimated from camera's outputs, but it is generally not the same case when a spoofing attack happens. Therefore, it is a good idea to exploit the coherence between these two sources of information as features for face spoofing detection. Our proposed solution relies mainly on this idea.

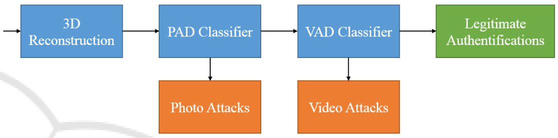


Figure 1: Flowchart of the whole proposed detection process.

The proposed solution consists of three major steps. Figure 1 shows the detection flowchart of the solution. Firstly, a 3D model of the face is estimated thanks to a 3D reconstruction process, section 3.1. Then, a Photo Attack Detection (PAD) classifier exploiting the 3D shape is employed to retrieve photo attacks (which use a static image of legitimate user, e.g. photo printed on paper or displayed on an LCD screen). The construction of PAD classifier is described in the section 3.2. For the ones which pass through the PAD classifier, they will be after that classified thanks to the Video Attack Detection (VAD) classifier, described in the section 3.3. The VAD classifier permits to detect video-replay attacks. The ones which finally pass through the VAD classifier would be considered as legitimate authentications.

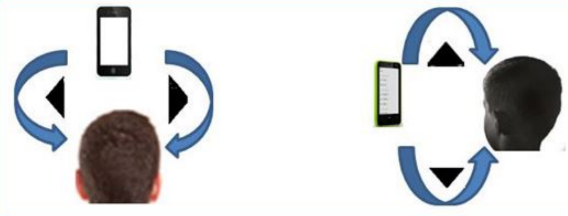


Figure 2: Camera movements during authentication process.

Apart from the necessary movement of smartphone (Figure 2) which requires the collaboration of user, all other process can be automatic. In this study, the video and sensors data collector, 3D reconstructor and classifier are regrouped inside a unique android application. However, the 3D reconstruction is not real-time yet that slows down the detection. In a real scenario, it is recommended to offshore 3D reconstruction and final classification to a dedicated server.

3.1 Facial 3D Reconstruction

In the process of proposed method, a three-dimensional model of the object (e.g. real face or fake face) is constructed from a video captured during the authentication process. For a better quality of the 3D model, the user is asked to move the phone's camera around their face in such a way as various head poses can be captured in the video. Two simple camera movements are considered in our proposed approach: in the vertical direction (i.e. upwards or downwards) and in the horizontal direction (i.e. to the left or to the right) (Figure 2). These proposed basic movements allow applications to easily communicate with users during authentication. They also permit to simplify the measure of coherence mentioned above.

In our study, 3D reconstruction process is assured by VisualSFM, a 3D reconstruction application developed by C. Wu (Wu, 2013) using Structure From Motion (SFM). The method requires a sequence of images in input. It gives as outputs the 3D reconstruction of the object captured as well as information related to the camera poses. Other recent solutions can replace VisualSFM in this step such as a faster 3D reconstruction proposed by Maninchedda et al. (Maninchedda et al., 2016). The choice of technology here does not affect the final outcome severely but highly depends on the computation capacity of smartphone.

The 3D model is a matrix which represents a cluster of featured points in a Cartesian coordinate system. Notes M the 3D model , it can be represented as:

$$M = \begin{bmatrix} x_1 & y_1 & z_1 \\ x_2 & y_2 & z_2 \\ \dots & \dots & \dots \\ x_N & y_N & z_N \end{bmatrix} \tag{1}$$

where N is the number of feature points and (x_k, y_k, z_k) $(k = 1, \dots, N)$ are the coordinates of the k-th point in the space $Oxyz$. Figure 3 gives an example of a real face reconstructed in form of points cloud.

The geometric information of i-th frame ($i = 1, .., n$) is also calculated and represent in position matrix: P_i and orientation matrix θ_i) $\P_i = \begin{bmatrix} x & y & z \end{bmatrix}$ and $\theta_i = \begin{bmatrix} \theta_x & \theta_y & \theta_z \end{bmatrix}$ where (x,y,z) is the coordination of the camera when it captured i-th frame $(\theta_x, \theta_y, \theta_z)$ is its orientation respectively in the view of Ox, Oy and Oz.

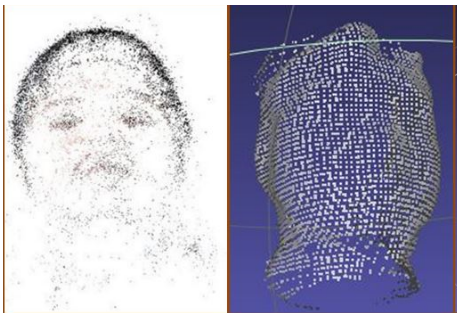


Figure 3: Points cloud of a real face 3D model.

3.2 Photo Attack Detection

In the case of photo attacks, the 3D reconstruction given by SFM method is clearly different from the one given in the case of a real face. Figure 4 shows different views of the 3D reconstruction of a printed face. It is easy to realize that the form of the 3D reconstruction is flattering in the case of photo attack. It can be explained by the fact that a real face is a real 3D object which contains much more depth information than a face printed on a piece of paper.

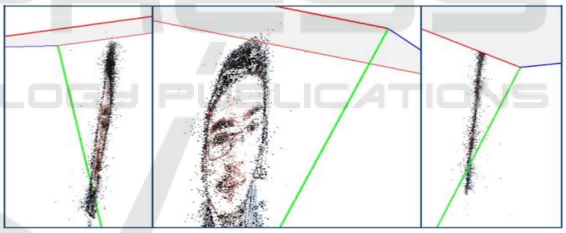


Figure 4: Different views of a printed face 3D model.

Thus, the more a 3D model is flat, the higher possibility of a photo attack. So that we can base on the thickness of the 3D reconstruction to eliminate photo attacks.

The thickness of 3D reconstruction can be relativity estimated using Principal Component Analysis (PCA). The PCA technique permits to transform the 3D reconstruction into a new coordinate system ($w = (w_{(1)}, w_{(2)}, w_{(3)})$), where each coordinate is represented by a principal component. This transformation is defined in such a way that the first principal component ($w_{(1)}$) has the largest possible variance (i.e. it accounts for as much of the variability in the data as possible), and each succeeding component, in turn, has the highest variance possible under the constraint that it is orthogonal to the preceding components. In that way, the variance of point cloud projected in the last component is the minimum among all vectors of space that can be used to represent the ”thickness”.

For a simplicity of the PCA transformation, the columns of the matrix M are firstly shifted to have a zero-mean. Without ambiguity, we use the same term M as the matrix shifted for the following development. The principal components matrix P is defined as an orthogonal linear transformation of the matrix M:

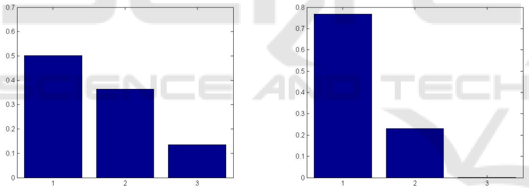
$$P = MW \tag{2}$$

where the matrix W is a 3-by-3 matrix whose columns are the eigenvectors of $M^T M$.

Denotes v_j the variance of the i-th column of P ($j = 1,2,3$). The order of magnitude of each column, denotes d_i, is given as follows:

$$d_i = \frac{v_i}{v_1 + v_2 + v_3} \tag{3}$$

For a face spoofing attack, the points of the 3D reconstruction is in a plan. Therefore, there is almost no information in the third component that makes d_3 very tiny. Meanwhile, for a real face, the thickness play a significant part in the total information. Figure 5 (a) and (b) give an illustration of the three principal components obtained respectively from fake and real face 3D models.



(a) (b)

Figure 5: PCA of real (a) and fake (b) face 3D reconstruction.

The different is so net that a simple SVM classifier fed by the order of magnitude d_i can be employed as a PAD classifier without any other processing.

3.3 Video Attack Detection

In the scenario of a video attack, a clip of authentic user’s head is displayed in a LCD screen in front of the camera. The video is edited so that the head moves in the same way to mimic the process. This attack can pass the PAD classifier since the moving head can provide different views of user’s face to construct a genuine 3D model of the head. The form of the 3D reconstruction does not give us enough information for spoofing detection. Therefore, we proposed to study, in addition, the camera poses.

In fact, the movement of the camera can be observed in two ways. The first is the positions and the poses of camera corresponding to each image. These positions which represent the movement of the camera according to the user’s face can be located by 3D reconstruction. Figure 6 shows an example of the camera positions estimated from the 3D reconstruction.

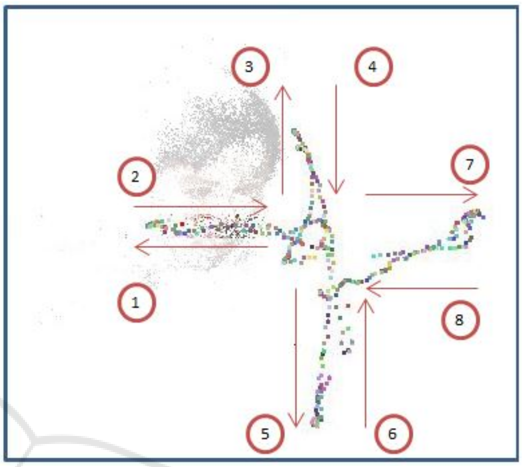


Figure 6: Example of the camera positions estimated from the 3D reconstruction. Direction of the camera move is marked form 1 to 8.

The 3D reconstruction describes the movement by position vectors P_i for the camera position and by orientation vectors θ_i for the camera poses with $i = 1,..,n$ is the index of the frame. Since the movement is mono-direction (the camera move by x-axis, y-axis but not both at the same time), the position and orientation can be represented separately axis by axis in the form of sequence:

$$X_i = P_{i,x} \tag{4}$$

$$\theta_i^x = \theta_{i,x} \tag{5}$$

The second observation is the trajectory of camera captured by movement sensors (e.g. accelerometer, gyroscope). These sensors observe the acceleration of translation and rotation of camera continuously by the time that allows to describe that movement independently. The gyroscope captures rotation acceleration and the accelerometer captures linear acceleration of the device affected by the force of gravity.

$$\frac{d^2x(t)}{dx^2} = a_x(t) \tag{6}$$

$$\frac{d^2\theta_x(t)}{dx^2} = g_x(t) \tag{7}$$

Where $t \in \mathbb{R}^+$ is the index of time, g_x, g_y, g_z are data from gyroscope and a_x, a_y, a_z are data from acceleration.

By using integration method, the position and orientation of thecamera can be calculated from the acceleration as the initial speed and initial position are both zeros. Let's t_{max} is the length of video, one frame is taken each $\Delta t = \frac{t_{max}}{n-1}$ second. Another sequence of the camera state can be obtained from the sensor data:

$$\widehat{X_i} = x(t_i) \tag{8}$$

$$\widehat{\theta_i^x} = \theta_x(t_i) \tag{9}$$

where $i = 1,..,n$, $t_1 = 0$, $t_n = t_{max}$, $t_i = \Delta t.(i-1)$

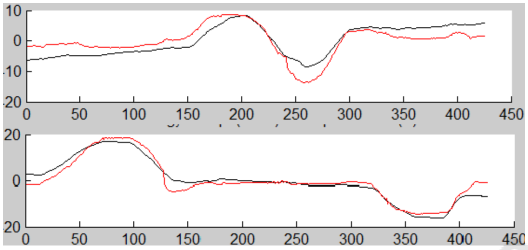


Figure 7: Correlations between θ_i^x (black)and $\widehat{\theta_i^x}$ (red) and between θ_i^y (black) and $\widehat{\theta_i^y}$ (red).

The camera poses are compared to the information captured from movement sensors (e.g. accelerometer, gyroscope). Both information sources (3D reconstruction and movement sensors) represent the movement of the camera according to the user's face. Thus, for a legit authentication case, there will be, a priori, a high similarity between these two sources. Meanwhile, when attackers use a video to authenticate, the camera poses estimated from the camera output will not represent the real poses of the camera. Therefore, the similarity between the two sources would not be assured. To estimate this similarity, a simple correlation can be applied for each couple of data as: (θ_i^x,$\widehat{\theta_i^x}$), (θ_i^y,$\widehat{\theta_i^y}$), (X_i,$\widehat{X_i}$) and (Y_i,$\widehat{Y_i}$). All these features (correlation results) are fed to an SVM classifier to form the VAD classifier mentioned previously.

4 EXPERIMENTAL RESULTS

The proposed video-replay attack detection process requires the data of motion sensors integrated within the smartphone. Actually, there are no public face spoofing datasets responding to this requirement. Therefore, we tested the proposed method in a specific database constructed in our laboratory.

The database includes 201 videos of 3 people including sensor data, therein: 99 cases of legitimate authentication, 63 cases of video-replay attack and 39 cases of photo attack.

The videos are captured in different light conditions and movement speed by three devices: 2 instances of Samsung Galaxy Alpha and a Samsung Galaxy Tab.

The database is divided into 2 sets. Each set has all three types of videos (legitimate, photo-attack and video-replay attack authentication). One set is used for training the PAD and VAD classifiers and the other set is used for testing. From the training set, we used just legitimate and photo attack authentication videos to train the PAD classifier and accordingly we used just legitimate and video-replay attack authentication videos to train the VAD one.

We also implemented the multi-scale Local Binary Pattern method (Maatta et al., 2011), which applies for a single image. Therefore, frames from the videos are used as input to implement this method.

Figure 8 shows the Receiver Operating Characteristics (ROC) curves of the proposed method and the LBP one. Our method performs much better than the LBP one especially for a small rate of false positive.

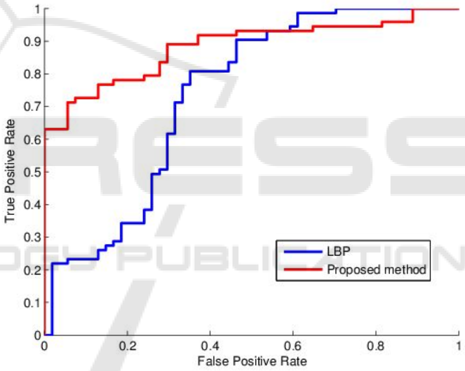


Figure 8: ROC curve for the proposed detection method in comparison with the one of LBP method.

5 CONCLUSION

In this paper, we propose a novel approach dedicated to mobile systems against the face spoofing detection problems. Our proposed approach is based on the 3D reconstruction form of face images to detect a photo attack. The detection of video attacks is based on the coherence between the movements captured by the smartphone sensors and the camera position estimated from 3D reconstruction.

Experimental results show that the approach eliminates attacks using a printed image efficiently. For video face spoofing attacks, the approach can still reach a high performance.

The 3D reconstruction process, which is realized with the help of the SFM method, needs a high computational cost. For future work, an improvement of this one will be envisaged. We plan to develop a real-time lightweight 3D reconstruction method dedicated to our face spoofing detection approach. We also think about a new strategy for frame selection in the 3D reconstruction process to reduce the computation time.

REFERENCES

Abdullah, M., Sayeed, M., Sonai Muthu, K., Bashier, H., Azman, A., and Ibrahim, S. (oct. 2014). Face recognition with symmetric local graph structure (SLGS). *Expert Systems with Applications*, 41(14):6131–6137.

Bao, W., Li, H., Li, N., and Jiang, W. (avr. 2009). A liveness detection method for face recognition based on optical flow field. In *International Conference on Image Analysis and Signal Processing*, pages 233–236.

Bashier, H., Lau, S., Han, P., Ping, L., and Li, C. (2014). Face spoofing detection using local graph structure. In *Proceedings of the 2014 International Conference on Computer, Communications and Information Technology*.

Housam, K., Lau, S., Pang, Y., Liew, Y., and Chiang, M. (2014). Face spoofing detection based on improved local graph structure. In *2014 International Conference on Information Science and Applications (ICISA)*, pages 1–4.

Hyung-Keun, J., Sung-Uk, J., and Jang-Hee, Y. (2006). Liveness detection for embedded face recognition system. *International Journal of Biological and Medical Sciences*.

Kim, G., Eum, S., Suhr, J., Kim, D., Park, K., and Kim, J. (mar. 2012). Face liveness detection based on texture and frequency analyses. In *2012 5th IAPR International Conference on Biometrics (ICB)*, pages 67–72.

Kim, S., Yu, S., Kim, K., Ban, Y., and Lee, S. (jun. 2013). Face liveness detection using variable focusing. In *2013 International Conference on Biometrics (ICB)*, pages 1–6.

Kollreider, K., Fronthaler, H., Faraj, M. I., and Bigun, J. (sep. 2007). Real-time face detection and motion analysis with application in liveness assessment. *Trans. Info. For. Sec.*, 2(3):548–558.

Maatta, J., Hadid, A., and Pietikainen, M. (oct. 2011). Face spoofing detection from single images using microtexture analysis. In *2011 International Joint Conference on Biometrics (IJCB)*, pages 1–7.

Maninchedda, F., Hne, C., Oswald, M. R., and Pollefeys, M. (2016). Face Reconstruction on Mobile Devices Using a Height Map Shape Model and Fast Regularization. In *2016 Fourth International Conference on 3D Vision (3DV)*, pages 489–498.

Nguyen, H. P., Retrainty, F., Morain-Nicolier, F., and Delahaies, A. (2016). Face spoofing attack detection based on the behavior of noises. In *2016 IEEE Global Conference on Signal and Information Processing, GlobalSIP 2016, December 79, Greater Washington, D.C., USA, 2016*, pages 1–5.

Sun, L., Pan, G., Wu, Z., and Lao, S. (2007). Blinking-based live face detection using conditional random fields. In *Proceedings of the 2007 International Conference on Advances in Biometrics*, ICB'07, pages 252–260. Springer-Verlag.

Wu, C. (2013). Towards linear-time incremental structure from motion. In *Proceedings of the 2013 International Conference on 3D Vision*, 3DV '13, pages 127–134. IEEE Computer Society.