A Countermeasure against Face-Spoofing Attacks Using an Interaction Video Framework

Kam Kong, Xiali Hei
Dept. of CIS
Delaware State University
Dover, DE 19901, USA
Email: {kkong, xhei}@desu.edu

Ting Zeng, Caijin Ling
School of EIE
Heyuan Polytechnic
Heyuan, Guangdong 517000, China
Email: zengtforever@163.com, ling8983@gmail.com

Chao Zhang
Dept. of Math
Delaware State University
Dover, DE 19901, USA
Email: zc50120@163.com

Binheng Song
Graduate School at Shenzhen
Tsinghua University
Shenzhen, 518055, China
Email: bsong@tsinghua.edu.cn

Hui Cao
Dept. of EE
Xi'an Jiaotong University
Xi'an, Shaanxi 710049, China
Email: huicao@mail.xitu.edu

Michael Peays
Dept. of CIS
Delaware State University
Dover, DE 19901, USA
Email: mpeays15@students.desu.edu

***Abstract*—With an increase in acceptance of face recognition systems, the desire for accurate biometric authentication - face recognition, has increased. Nowadays, the fundamental limitations of existing systems are, the vulnerabilities of false verification via a picture or simple video of a person. In this paper, it is inspired by how humans can perform reliable spoofing detection only based on the available scene and context information. We propose our framework in a combination of interaction model and Moiré pattern analysis to make sure that a user is a real person. It is different from some systems that use hardware-based liveness detection. We focus on the software-based approaches, in particular, the necessary algorithms that allow for a liveness detection in real-time. Our experiments results show excellent performance to the state of the art.**

***Index Terms*—Face Recognition; Spoofing Attacks; Liveness Detection; Moving Object Tracking; Moiré Pattern Analysis;**

I. Introduction

Facial recognition has achieved a great success during the past decades. However, they are also attracting attention from attackers. It is now increasingly aware that existing facial recognition systems are susceptible to forged face attacks. Unauthorized attackers try to access illegal authorities by exhibiting forged faces of an authorized client. Serious consequences may occur if these attacks succeed. Thus, robust and sufficient anti-spoofing techniques are needed.

Nowadays attackers can obtain a client's face images by using portable digital cameras or directly downloading them from the Internet. Printing photos or showing videos are most in commonly used in the trick because forged faces like photos and video playbacks are not only easy to implement but also usually quite efficient in spoofing a face recognitions system. On the other hand, the wax images and human-like robots are used by attackers artificially. It is difficult to prevent such attacks using existing approaches.

For our research/experiments, we hold the opinion that a perfect detection method of liveness should meet the following conditions:

- The process is robust and works in whatever conditions;
- The calculation process should be fast;
- It does not need more facilities;
- It can be a standalone model;
- The characteristic can be cultured easily.

We propose a framework based on "movement" interaction feedback to prevent the print attack, replay attack, wax image fraud attack, and human-like robot deception. It is different from the systems that use hardware-based liveness detection. We focus on the software-based approaches, in particular, the necessary algorithms that allow for a liveness detection in real-time. Inspired by how humans can perform reliable spoofing detection only based on the available scene and context information, we propose our framework in a combination of interaction model and Moiré pattern analysis to make sure that a user is a real person. In our method, we firstly use Moiré pattern analysis to prevent a replay attack; then, we judge if the face is a real person with the help of an interaction model. The interaction model is an important part of our framework. In this model we defined: interaction, activity, and threshold definition, motion direction, and the number of frames. The results show that the proposed framework could achieve robust and good results in the well-defined conditions. It is a lightweight and standalone model.

Our contributions are summarized as follows:

- We present a practical framework to prevent the spoofing attacks.
- Our method is robust in face recognition.
- To the best of our knowledge, we are the first group who proposed to a promising method in preventing the wax image fraud and human-like robot deceived.

Our framework is effective: (1) We get a success rate higher than 98% in our experiments; (2) Our solution is light-weight and less expensive, in which only an ordinary smart-phone is required; (3) Our solution can be widely applied to protect

978-1-5090-5363-6/17/$31.00 ©2017 IEEE

mobile devices and other credentials.

II. Our Proposed Framework

It is our assertion that the performance of spoofing detection techniques can be improved by leveraging moving objects. As shown in Figure 1, the proposed framework first performs video identification, then conducts the interaction model and carries out the judge model at the end. The paper mainly introduces the last two in our framework due to page limit.

Fig. 1. The proposed framework

A. An interaction Fundamental Model

We assume a real person can interact with a program well. If the user to be verified cannot follow several instructions consecutively, we think it is a video instead of a real person. This is the main assumption of our interaction model.

To register the frames based on background correlation, we consider the first frame as the reference and calculate context association for the subsequent.

1) Instruction Definition: We assume the video is a real one, so the activities in the video may be meaningful. However, the machine cannot recognize the interactive motions. Therefore, we should define and indicate what actions the user should take.

To generate instructions, we make a question set `A`, which contains many items, for instance:

- `a[0]="stand at attention, then stretch your left hand to the left" ;`
- `a[1]="hand up your left hand" ;`
- `a[2]="stand at attention, then stretch your right hand to the right" ;`
- `a[3]="hand up your right hand" ;`
- `a[4]="turn the head to the left" ;`
- `a[5]="turn the head to the left"` , etc.

The user needs to follow the machine's instructions, which are random from the program, and make the right activities.

2) Activity Definition: How does the machine know, and judge what activities the user do? We should make the machine understand the actions. To achieve this goal, first, we define what these activities are.

All movements in our framework are recognized activities, however, only these which are smaller than the threshold are considered as productive activities. On the other hand, only the activities which are corresponding with the instructions are right activities. We give those definitions of events corresponding to instructions.

To let the machine understand movements automatically, we make several assumptions for the video as follows:

- The valid user can follow the instruction, and only do what he/she is asked. Additional actions are not allowed;
- The extra movements of user are greater than threshold;
- The surroundings is fixed relative. There is no any interference motion in the video.

When we get a real video from a machine, we get some of the fixed time sequence frames. Assume the number of frames is N, $(N \geq 3)$.

To effectively track and present movement, we employ the sliding window approach with two frames. The previous frame is used for making a difference from the current frame. The sliding window is propagated until N frames by incrementing one frame at a time. The sliding window is used to detect the rate of the variation between two frames. As illustrated in the Figure 2, for kth frame, the differential information from the last previous frame is given by:

$$D(k) = abs(f(k) - f(k-1)) \tag{1}$$

where k, $(0 < k < N)$, $D(k)$ represents the variation between k^{th} frame and $(k-1)^{th}$ frame, and $f(k), f(k-1)$ denotes k^{th} frame and $(k-1)^{th}$ frame, respectively. Let $f(0)$ be a photograph before the interaction. Some instances are shown as follows: $D(1) = abs(f(1) - f(0))$;

$D(2) = abs(f(2) - f(1))$;

$D(3) = abs(f(3) - f(2))$;

$\vdots$

$D(N) = abs(f(N) - f(N-1))$;

Since $f(k), f(k-1)$ represent a certain photograph, without doubt $D(k)$ is also a picture. We can use a two-dimensional array to express the new photo $D(k)$. The element $d(i, j)$ of the two-dimensional array denotes the pixel of coordinate (i, j). In order to investigate the difference of the original pictures $f(k)$ and $f(k-1)$ clearly, we set a threshold Th to filter some weak target. The equation of $d(i, j)$ in threshold is shown as equation 2.

$$d(i,j) = \begin{cases} 1, d(i,j) < Th \\ 0, d(i,j) \geq Th \end{cases} \tag{2}$$

In the case of monochrome images, there are usually 256 gray levels. In some digital image processing applications, the number of levels is even decreased. By filtering of the threshold, the $D(k)$ is converted to black and white color, which are shown as figure 2. Providing information on the image consisting only two levels, thus, it will be a binary image. Therefore, the machine can realize it apace and precisely.

Moreover, we also should predict the track of movement in sixteen districts indicator model (SDIM), which is shown in Fig. 3 and would be explained in our Judgment Model.

3) Threshold Determination: We use the threshold in above subsection. It will affect the effectiveness directly. Therefore, an appropriate value of the threshold is crucial in our experiment. However, the best value is not always proper. After numbers of operations exploration and unremitting efforts, we have found the right range of threshold which is between fifty to seventy.

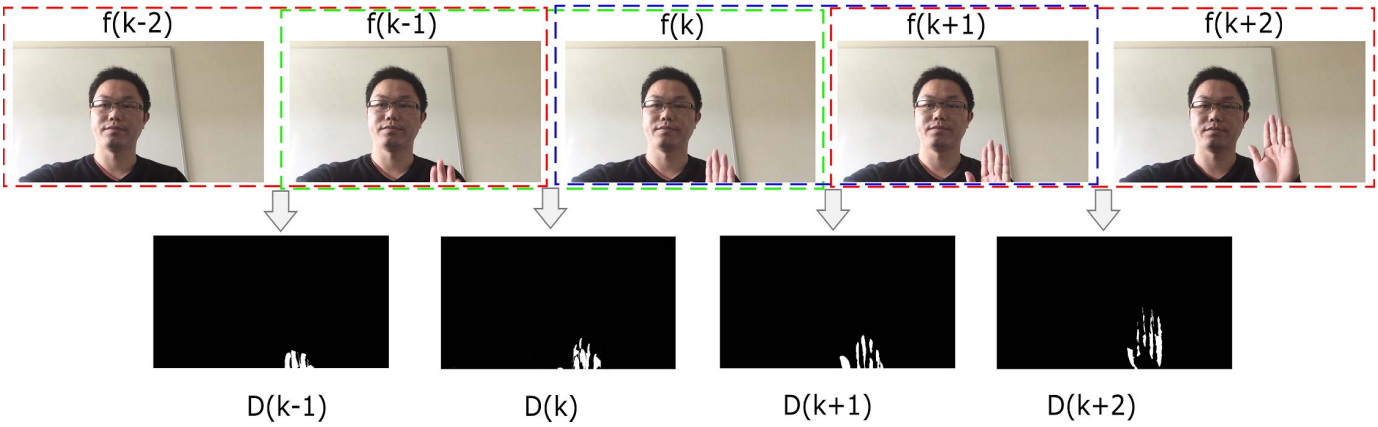

Fig. 2. Schematic of slide

4) Motion Direction: To enhance security, after we get the frames in our framework, we employ random frames. However, if the frames got are not in a sequence, we cannot judge the direction of movement. Also, the interval between two frames has a great impact in direction. Therefore, our random algorithm needs to consider both of them. *Notably, it cannot recognize the target, when the number of frames is minuteness.*

5) Number of Frames: The number of video frames will affect results. The number of frames actually tracks movement object. Indeed, it would spend more time to that if get more frames to judge the action. Three is the minimum of the frames, as Figure 2 showing. Thus, it can produce two value of $D(k)$ for judgment of moving an object. However, more values of $D(k)$ can make a better and more sensitively reflecting movement track. The biggest number of frames can not be greater than the ratio of the frame-rate of video and the duration of the video. In other words, the range of value N must satisfy Equation 3.

$$\begin{cases} 3 \le N \le N_{max} \\ N_{max} = \frac{video.Duration}{video.FrameRate} \end{cases} \tag{3}$$

B. Decision Making Model

To meet more requirements, we build a sixteen position indicator model relative user like Fig. 3. Assume that we can obtain the frames, which are distributed in different districts of Fig. 3. And making an assumption of that there is a track between two frames. Thus, the machine combining the defined activities can recognize the movement.

1) The Sixteen Districts Indicator Model (SDIM): Using this automatic model, we can judge the position of the action in a frame. We define the sixteen districts according to the structure of the people. First, we set up a center line base on the body of the user. Then, we build two vertical lines next to the hand `a` units, which equals to the size of the head of the user. On the other side, we establish the fist horizontal line through the eyes of the user; then we set up the second and third horizontal lines via the position of hands. That means a sixteen districts of a picture made by three vertical lines and three horizontal lines.

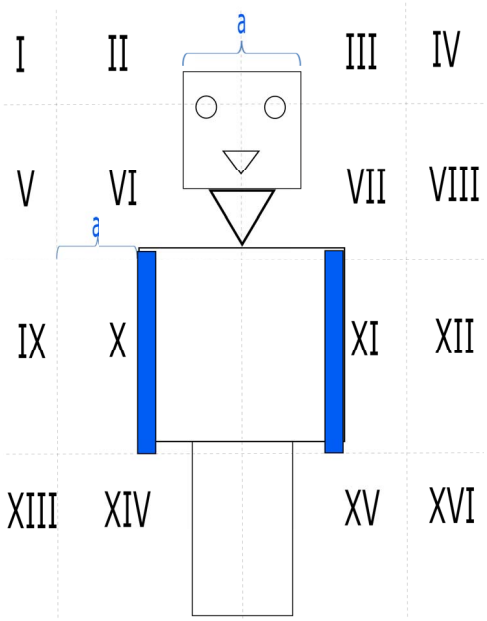

Fig. 3. SDIM

However, it is not necessary having sixteen sections sometimes. The specific requirement is up to the instruction and activities definition.

2) Decision Making: Fig. 4 shows the process of decision. We can see the target is in the section V from Fig. 4(a), and another target is located in section II of SDIM as showing Fig. 4(b). Therefore, it is easy to judge whether the target is moving from section V to III. Thus, combining the activities definition and SDIM, the machine can make a right judgment for the user.

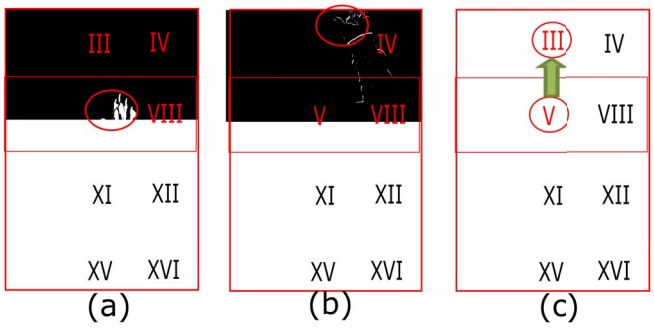

Fig. 4. An example of SDIM

C. The Video Identification

To insure the video captured by face recognition machine is a real video, a high-efficiency identification method should be adopted. We employ Moiré pattern analysis and figure magnification to verify the video. The display of digital devices (laptops, mobile devices, and tablets) exhibit a naturally occurring fixed repetitive pattern created by the geometry of color elements that are used for color displays. Therefore, whenever a video of a digital screen is recorded, Moiré Patterns will naturally present themselves.

Inspired by the mechanism of a display, we propose this method is searching Moiré patterns due to the overlap of the digital grids. The conditions under which these patterns arise are described, and the proposed detection is based on peak detection in the frequency domain. We can detect the video whether it is a replay video. Literature [1]–[6] introduce Moiré-Pattern and its usages [7]. Based on large numbers of experiments, we are sure that it is a robust method to identify a replay video. Fig. 5 shows the Moiré Pattern produced by three different display of the camera. Fig. 5(a) is a frame from a replay video in a computer which is recording by an iPhone 4S; Fig. 5(b) is a frame from a replay video in a computer which is recording by an iPad min 4; It is not evident to see the Moiré Pattern by eyes in 5(c), which is a frame from a replay video in an iPad min 4 recorded by an iPhone 4S. However, after employing Difference-of-Gaussians (DoG) filter in [7], [8], we can find the Moiré Pattern easily.

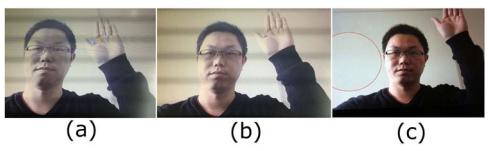

Fig. 5. Video identification process

Nevertheless, employing only Moiré Pattern analysis and figure magnification can not discover robots and wax figures. We will discuss the techniques to prevent such attacks in the discussion section.

III. Experiments

Two sets of experiments were carried out. The first one investigated the sensitivity of Moiré Pattern analysis method to check the video whether it is valid. The second one verified the ability to use the proposed judgment model in section III to detect the track of a moving object. The cameras we used to take all videos are in an Ipad mini 4 and an iPhone 4S. All the experiments were done in MATLAB R2016a under Professional of Windows 7. The hardware of circumstance is shown in Table I.

TABLE I
THE HARDWARE OF CIRCUMSTANCES

Hardwares	Configurations
Processor	i7-4810MQ CPU @ 2.80GHz
Installed memory (RAM)	16.0 GB(15.6 GB usable)
System type	64-bit Operating System
camera	iPad min 4, and iPhone 4S

A. Videos Preparation

To let the experiments more widely and significant, we took 160 videos, which includes 80 real-access videos and 80 replay-attack videos. We also use the CASIA's public database. Table II shows the details.

TABLE II
THE VIDEOS FOR EXPERIMENTS

Device	No. of Videos	Type	From
iPad min 4	40	real	our team
iPhone 4S	40	real	our team
iPad min 4	40	spoof	our team
iPhone 4S	40	spoof	our team
iPad	100	real	CASIA
iPad	100	spoof	CASIA

B. Video Checking

To make sure the video took from a real user, firstly, we should prevent a replay attack. As the description in section III, we use the Moiré pattern analysis to check the video. Three steps can finish the identification.

- Step 1. Randomly get a framed photo from the video.
- Step 2. Magnify the picture got from Step 1.
- Step 3. Using the Moir pattern analysis technique to check the photo.

If it can find the Moiré pattern in Step 3, the video is a fake video. We did the experiment with 360 videos, including ten real videos and ten fake videos. Table III shows the results of our experiments.

TABLE III
DATA OF THE VIDEOS IDENTITY OF EXPERIMENTS

Device	No. of Videos	Type	From	No. of moire
iPad min 4	40	real	our team	0
iPhone 4S	40	real	our team	0
iPad min 4	40	spoof	our team	40
iPhone 4S	40	spoof	our team	40
iPad	100	real	CASIA	0
iPad	100	spoof	CASIA	96

C. Judge the Movement

Because the videos which are from CASIA do not match conditions our solution needs, thus, we did this experiment with the remaining 120 videos. We adopt six frames in Interaction Fundamental Model, and the value of the threshold is 66. The results are shown in Table IV, where FFR means False Rejection Rate, and FAR says False Acceptance Rate (the definition would discuss in next section).

TABLE IV
JUDGE OF EXPERIMENTS

Device	No. of Videos	Type	FFR	FAR
iPad min 4	40	real	0	0
iPhone 4S	40	real	2	0
iPad min 4	40	spoof	0	2
iPhone 4S	40	spoof	0	0

IV. PERFORMANCES EVALUATIONS

Half Total Error Rate (HTER) is defined as the half of the sum of the False Rejection Rate (FRR) and False Acceptance Rate (FAR). These which define as Equation 4 is used to measure the performance on our proposed framework. These parameters are used to measure the performance of face recognition.

$$\begin{cases} HTER = \frac{FRR+FAR}{2} \\ FRR = \frac{FR}{NI} \\ FAR = \frac{FA}{NR} \end{cases} \tag{4}$$

Where, FR and NI denote False Rejection, and Number of Imposter; FA, NR is short for False Acceptance, and Number of Real, respectively.

According to the Equation 4, FRR = $\frac{FR}{NI} = \frac{2}{160}, FAR = \frac{FA}{NR} = \frac{1}{80}, therefore, HTER = \frac{FRR+FAR}{2} = \frac{1}{80} = 1.25\%$. Combining the success rate of video in identify, the HTER = $1.25\% * \frac{1}{1-4/360} = 1.26\%$. The best performance in [9] on the CASIA dataset is a HTER of 21.01%, which is much higher than our error rates.

V. DISCUSSIONS

The results show that our proposed framework can significantly distinguish a real user, the real-access videos, and replay-attack videos. Moreover, our framework also can tell the difference between the real-access video and robot's video when we add time as a parameter. Since we can get the variation of a user in time, and assume that user is continuously and uniformly moving, thus, it is a constant in the same interval.

Whereas, the robot cannot conduct this like a human. Different motors make all his movements. Therefore, the motion is not continuous and uniform as well as a real user. Fig. 6 shows the different relationship between **s** and time in a real user and a robot. Fig. 6(a) denotes a real user, and Fig. 6(b) is a robot. Therefore, the machine can decide the video is from

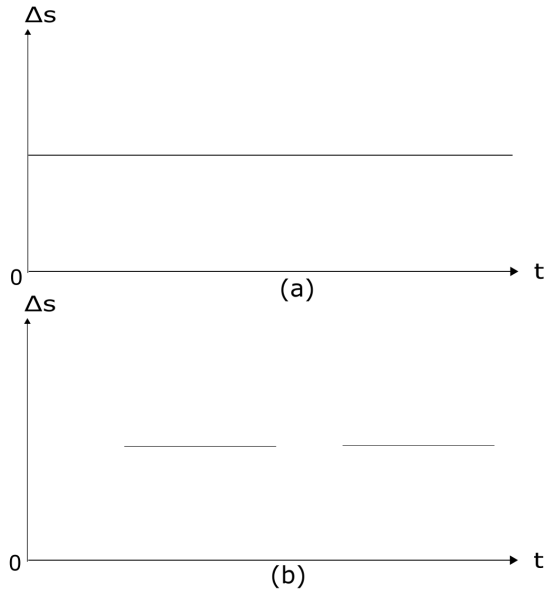

Fig. 6. Δs-t relationship

a real user or a humanoid robot.

VI. RELATED WORK

Motion-based approaches can be further divided into two categories: physiological responses based and physical motion based. The physiological responses, such as eye blinking, mouth (lip) movement, are important clues to verify the "liveness". Pan et al. used eye blinking for face anti-spoofing [10], [11]. In their method, a conditional random field was constructed to model different stages of eye blinking. In [12], Kollreider et al. used lip movement classification and lip-reading for the purpose of liveness detection. Furthermore, Chetty et al. [13], [14] proposed a multi-modal approach to aggrandize the difficulty of spoofing attacks. It determined the liveness by verifying the fitness between video and audio signals.

The works in [11], [15] bring a real-time liveness detection specifically against photo-spoofing using eye-blinks which are supposed to occur once every 2-4 seconds in humans. The system developed uses an undirected conditional random field framework to model the eye-blinking that relaxes the independence assumption of generative modeling and state

dependence limitations from hidden Markov modeling. With this setup, the proposed detector can achieve 95.7% true-positive classification against a false alarm of less than 0.1% when considering a simultaneous blink of both eyelids in all test samples. However, as a biological organ, the human eye is prone to flash at regular intervals. These intervals or frequency of blinking varies individually from person to person. Also, the eye-blinks can be forged when using print attacks and make a hole in the eyes areas.

Face spoofing detection methods based on face analysis extract face-specific characteristics (physiological or behavioral) such as eye blink [11], [15], lip or head movement [16], texture [17], and 3D shape [18], [19]. Some methods used a fusion of multiple physiological or behavioral clues to detect spoof faces [16]. Although these methods report favorable results for intra-database testing, they require an accurate face and/or landmark (eye) detection.

VII. Conclusion

Anti-spoofing in face recognition systems must be quickly mature to provide a robust and computationally efficient solution and improve the practicality of face recognition system. Our research presents a practical framework for spoofing detection in face recognition systems. Using an interactive model and moire pattern analysis in face recognition, the machine can automatically recognize a video, then make sure the interactor is a human. Our experiments indicate that our proposed framework is a meaningful solution of liveness detection. Further, we present a human motion estimation based technique using an interactive model. Our framework provides high performance on the replay attack and humanoid robot attack. We believe our scheme can be further improved through incorporation of machine learning later.

Acknowledgments

This work was supported in part by the State of Delaware Federal Research and Development Matching Grant Program (DEDO start-up grant) and US NSF under grants CNS-1566166.

References

[1] J. Allebach and B. Liu, "Analysis of halftone dot profile and aliasing in the discrete binary representation of images*," *JOSA*, vol. 67, no. 9, pp. 1147–1154, 1977.

[2] M. Takeda, H. Ina, and S. Kobayashi, "Fourier-transform method of fringe-pattern analysis for computer-based topography and interferometry," *JosA*, vol. 72, no. 1, pp. 156–160, 1982.

[3] J. C. Krumm and S. A. Shafer, "Sampled-grating and crossed-grating models of moire patterns from digital imaging," *Optical Engineering*, vol. 30, no. 2, pp. 195–206, 1991.

[4] J. Leendertz and J. Butters, "An image-shearing speckle-pattern interferometer for measuring bending moments," *Journal of Physics E: Scientific Instruments*, vol. 6, no. 11, p. 1107, 1973.

[5] Z. Y. Rong and P. Kuiper, "Electronic effects in scanning tunneling microscopy: Moiré pattern on a graphite surface," *Physical Review B*, vol. 48, no. 23, p. 17427, 1993.

[6] I. Amidror, I. Amidror, I. Amidror, and I. Amidror, *The theory of the moiré phenomenon*. Springer, 2000, no. LSP-BOOK-2000-001.

[7] D. Caetano Garcia and R. L. de Queiroz, "Face-spoofing 2d-detection based on moiré-pattern analysis," *Information Forensics and Security, IEEE Transactions on*, vol. 10, no. 4, pp. 778–786, 2015.

[8] Z. Zhang, J. Yan, S. Liu, Z. Lei, D. Yi, and S. Z. Li, "A face antispoofing database with diverse attacks," in *Biometrics (ICB), 2012 5th IAPR international conference on*. IEEE, 2012, pp. 26–31.

[9] I. Chingovska, A. Anjos, and S. Marcel, "On the effectiveness of local binary patterns in face anti-spoofing," in *Biometrics Special Interest Group (BIOSIG), 2012 BIOSIG-Proceedings of the International Conference of the*. IEEE, 2012, pp. 1–7.

[10] L. Sun, G. Pan, Z. Wu, and S. Lao, "Blinking-based live face detection using conditional random fields," in *Advances in Biometrics*. Springer, 2007, pp. 252–260.

[11] G. Pan, L. Sun, Z. Wu, and S. Lao, "Eyeblink-based anti-spoofing in face recognition from a generic webcamera," in *Computer Vision, 2007. ICCV 2007. IEEE 11th International Conference on*. IEEE, 2007, pp. 1–8.

[12] K. Kollreider, H. Fronthaler, M. I. Faraj, and J. Bigun, "Real-time face detection and motion analysis with application in liveness assessment," *Information Forensics and Security, IEEE Transactions on*, vol. 2, no. 3, pp. 548–558, 2007.

[13] G. Chetty and M. Wagner, "Audio-visual multimodal fusion for biometric person authentication and liveness verification," in *Proceedings of the 2005 NICTA-HCSNet Multimodal User Interaction Workshop-Volume 57*. Australian Computer Society, Inc., 2006, pp. 17–24.

[14] G. Chetty, "Biometric liveness checking using multimodal fuzzy fusion," in *Fuzzy Systems (FUZZ), 2010 IEEE International Conference on*. IEEE, 2010, pp. 1–8.

[15] G. Pan, L. Sun, and Z. Wu, *Liveness detection for face recognition*. INTECH Open Access Publisher, 2008.

[16] S. Bharadwaj, T. Dhamecha, M. Vatsa, and R. Singh, "Computationally efficient face spoofing detection with motion magnification," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops*, 2013, pp. 105–110.

[17] J. Määttä, A. Hadid, and M. Pietikainen, "Face spoofing detection from single images using micro-texture analysis," in *Biometrics (IJCB), 2011 international joint conference on*. IEEE, 2011, pp. 1–7.

[18] W. Bao, H. Li, N. Li, and W. Jiang, "A liveness detection method for face recognition based on optical flow field," in *Image Analysis and Signal Processing, 2009. IASP 2009. International Conference on*. IEEE, 2009, pp. 233–236.

[19] M. De Marsico, M. Nappi, D. Riccio, and J.-L. Dugelay, "Moving face spoofing detection via 3d projective invariants," in *Biometrics (ICB), 2012 5th IAPR International Conference on*. IEEE, 2012, pp. 73–78.