Designing a Facial Spoofing Database for Processed Image Attacks

Luma Omar and Ioannis Ivrissimtzis

Durham University
School of Eng. & Comp. Sciences
South Road, Durham, DH1 3LE, UK
{luma.omar, ioannis.ivrissimtzis}@durham.ac.uk

Keywords: Face recognition; face anti-spoofing; face image database; processed image imposter attack

Abstract

Face recognition systems are used for user authentication in everyday applications such as logging into a laptop or smartphone without need to memorize a password. However, they are still vulnerable to spoofing attacks, as for example when an imposter gains access to a system by holding a printed photo of the rightful user in front of the camera. In this paper we are concerned with the design of face image databases for evaluating the performance of anti-spoofing algorithms against such attacks. We present a new database, supporting testing against an enhancement of the attack, where the imposter processes the stolen image before printing it. By testing a standard anti-spoofing algorithm on the new database we show a significant decrease in its performance and, as a simple remedy to this problem, we propose the inclusion of processed imposter images into the training set.

1 Introduction

Face recognition systems offer a fast, reliable, convenient and inexpensive way for person authentication. However, the relatively easy access to face image data of the average person, for example by downloading photos or videos they have posted on social media sites, means that face recognition is particularly vulnerable to spoofing attacks. As a result, its use is restricted to either applications where security is considered secondary to convenience, e.g. log in to a personal laptop or a smartphone, or to applications in tightly controlled environments, such as passport control in airports. When either of these two conditions is not met, e.g. money withdrawals from a street ATM machine, face recognition is not deemed a suitable person authentication method.

In one of the simplest spoofing attacks on a face recognition system, an imposter might print on paper the photo of the rightful user and hold it in front of the camera of the face recognition system to gain access. Despite the lack of sophistication of such an attack, it is nevertheless capable to gain access to a laptop or a smartphone protected by some of the most popular face recognition user authentication systems [1]. As a response to such vulnerabilities, the development of anti-spoofing algorithms and techniques, commonly called *liveness tests*, has become a very active research area.

The performance of anti-spoofing algorithms is evaluated on databases containing both photos of real people called *client images*, and photos of imposters, which essentially are photos of client images and are called *imposter images*. The design of such a database is a particularly challenging task given the multiple sources of variation in spoofing attacks. Indeed, a whole range of choices, from the choice between a paper photo and an electronic display for the attack, to the type of paper and printer used to print a photo, to the size of that photo and the way it is held in front of the camera, all these factors can impact the effectiveness of the attack and thus the perceived performance of the anti-spoofing algorithm.

In this paper we address a gap in the current practice of the anti-spoofing algorithm evaluation, namely the assumption that the attacker prints the photo of the rightful user as it is, i.e. without attempting to process it in order to increase the effectiveness of the attack. We created a face image database which besides the usual imposter images it also contains imposter images obtained by photo-shooting printouts of sharpened client images, see Figure 1. We thested the database on a standard liveness test [2] and found that the more sophisticated attack with processed imposter images is more likely to evade detection.

Contribution: the main contribution of the paper is the design and construction of a database of face images for testing anti-spoofing algorithms, which, to the best of our knowledge, is the first one based on the assumption that the imposters may use image processing tools to enhance the effectiveness of their attack.

Limitation: as the main limitation of the paper we note that our current database only serves as a proof of concept, considering only one image processing operation on the client images before they are printed. However, extending the database with imposter images that have undergone other types of processing is a relatively straightforward, even though laborious, process.

The rest of the paper is organised as follows. In Section 2 we review the relevant literature. In Section 3 we describe the details of the database. In Section 4 we briefly describe the liveness test we use for testing. In Section 5 we test the database by running on it a standard liveness test and discuss the relevance of the results. We conclude in Section 6.

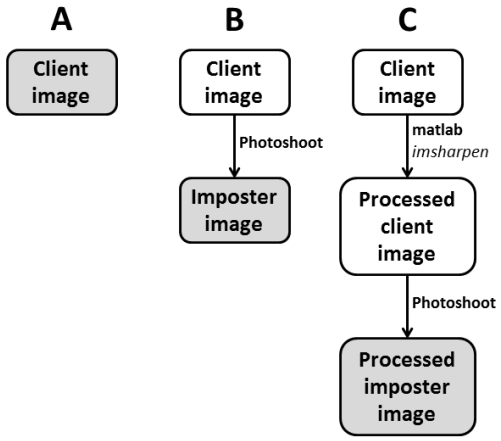

Figure 1. The standard database design consisting of client (A) and imposter images (B), is augmented with processed imposter images (C).

2 Related work

Face recognition is a well established research area with the state of the art techniques achieving recognition rates that rival the human ability to recognize faces under similar conditions. The input of a face recognition algorithm can be a grayscale or a color still image, a short video sequence or a 3D scan of someones face. The first examples of successful face recognition algorithms from still images were based on Principal Component Analysis (PCA) [3] and [4, 5], while further improvements proposed in [6] were able to cope with large scale databases and handle better the problem of pose variability by using modular eigenspaces.

Biometric attacks fall in two main categories: direct and indirect. Direct attacks are based on the use of stolen biometric information, either in digital form such as digital audio, images or video, or in physical form, for example a fingerprint on a gelatinous membrane. Indirect attacks rely on the use of computational algorithms to construct biometric information and use it to access the security system. Galbally et al. tested face recognition systems based on PCA [7] and GMM and PCA [8] against hill-climbing indirect attacks. In all cases it was found that the attack could spoof the systems, but the GMM was found more robust.

Liveness tests aim at distinguishing between live faces in front of the camera and imposters. In the last few years they have developed rapidly [9, 10, 11], although it is clear that they can not be considered a mature technology yet. Head or face movement and eye-blinking detection methods have been successfully used liveness tests. Li et al. [12] proposed a liveness test based on the different distributions in the frequency domain of light reflected from a flat 2D surface or a 3D surface. Pan et al. [13] use eye-blinking detection for a liveness test based on a probabilistic model of eye-blinking behaviour. In [14] optical flow lines are used to model and distinguish between live face movement and still image movement. Lai and Tai [15] recently proposed a liveness test against attacks by fake images or videos displayed on HD screens by analyzing the chrominance characteristics and the saturation of the face recognition system's input images. Convolutional neural networks have also been recently employed to assist anti-spoofing techniques as in [16] where the classifier is trained to distinguish between live faces and imposter images from their differences in the edge patterns and the surface textures in single frames of video sequences.

The liveness test we used to evaluate the Durham Face Database was proposed by Tan et al. in [2]. There, they proposed several variants of the basic algorithm and the one we chose here trains a sparse logistic regression classifier with the difference of Gaussians of the database images. In a further improvement to the Tan et al. algorithms, [17] apply contrast-limited adaptive histogram equalization before computing differences of Gaussians, increasing the robustness of the test under bad illumination conditions. Differences of Gaussians are also used in [18], where a Support Vector Machine is trained.

The idea that a still image attack can be enhanced by digitally processing the face image before printing it had been suggested by the authors in [19]. However, as there are no publicly available face image databases containing processed imposter images as in Figure 1(C), we could not test our assumption. Instead, we used the publicly available NUAA database, measured the performance of the liveness test in [2] against digitally sharpened versions of imposter images, see Figure 1(B), and argued that the drop in the performance of the liveness test as evidence supporting our assumption.

2.1 Databases

The design of a database for evaluating liveness tests is a challenging task, given the variability of form of the imposter attacks. An additional difficulty is that the performance of the current state-of-the-art liveness tests seems to drop significantly when the imposter attack deviates even slightly from the protocol that was used to produce the imposter samples of the training set. That is, for example, when a different paper type is used, or a different printer or electronic display, or camera. A similar problem has been observed in the behaviour of state-of-the-art algorithms for the classic face recognition problem, where it is usually referred to as the *interoperability* problem, see for example Gallbaly and Satta [20]. As a result, it is important to have an exact description of the protocol under which a face image database was constructed, even if that contains a number of tedious and seemingly irrelevant details.

Next, we review the NUAA database, which as the Durham Face Database consists of still images, and three databases containing short video sequences, which nevertheless can also be used to evaluate still image liveness tests after extracting frames from the video sequences. We note that none of these four databases contains images or videos produced by processed imposter images, see Figure 1(C).

NUAA is a publicly avaliable database [2] containing face images of 15 participating people, taken under non-uniform lighting conditions. Participants were from both genders and were asked to assume neutral facial expressions without head movement or eye-blinking. The client image part of the database,

Figure 1(A), consists of 500 coloured images of each participant, with 640×480 resolution, taken by a conventional web-camera with a frame rate of 20fps. Client images were printed in three different sizes, 6.8cm$\times$10.2cm and 8.9cm$\times$12.7cm on photographic paper and 70g A4 paper, using an HP colour printer. Imposters images, Figure 1(B), were produced from these printouts with a Canon camera from a distance that would allow the face to cover approximately 2/3 of the whole scene.

PRINT-ATTACK is a publicly available database [21] containing videos from 50 participants. The videos were captured under two different conditions: controlled and adverse. In the controlled environment the background lighting conditions are uniform, while the adverse environment has uncontrolled background lighting conditions. The database consists of a total of 200 real access videos and 200 videos of spoofing attempts, which used A4 printed photos of the participants.

REPLAY-ATTACK was presented in [22] as an attempt to enrich and overcome shortcomings of the PRINT-ATTACK database. It contains short video recordings from 50 participants. Real-access videos were captured with an acquisition system built on an Apple 13-inch MacBook laptop. They consist of a couple of 15-second video clips for each participant with 320$\times$240 resolution at 25fps. For the spoofing attempts, two high resolution shots of each person using a 12.1 megapixel Canon PowerShot SX150 IS camera and an iPhone 3GS 3.1 mega pixel camera were taken. Spoofing attempts were classified according to three different conditions: (i) photos printed on A4 paper (ii) iPhone mobile display video and photos playbacks (iii) photos and videos been displayed on an iPad screen with 1024$\times$768 resolution. The REPLAY-ATTACK database was then used to evaluate the performance of a liveness test based on histograms of Local Binary Patterns, as proposed in [23], and a support vector machine as classifier.

CASIA is a database presented [18] and it was designed with the maximisation of the variability as its main aim. The client images come in three different imaging qualities: low, normal and high. The spoofing attempts made use of either printed photos on copper paper, or an iPad display. The database consists of 12 videos for each subject, 3 of which are genuine and 9 are spoofs.

3 Database

The Durham Face Database contains face images from 21 people. All photo-shooting sessions took place in the Imaging Laboratory of Durham University and for each participant a total of 50 photos were taken using a professional Canon EOS Rebel T3i (600D) with a 18-250mm lens. The original photos, together with the corresponding cropped images, the imposter images and the processed imposter images produced by them are all publicly available at Durham Face Database.

Client images: The camera was mounted on a tripod and operated with the default autofocus settings at $5,184 \times 3,456$ resolution. To isolate as much as possible the effect of the image sharpening that w applied to create the processed imposter images, all client images were taken in frontal view, with neutral expressions, under uniform illumination and background conditions. The raw client images were cropped down to size 640×640, which gives a good balance between image quality and the speed of training and testing the classifier. For our purposes, it was important to avoid any resizing of the images, since that would mean an extra image processing operation with a largely unpredictable and difficult to account for effect. To achieve this, all participants were seated between 1m and 1.25m away from the camera, at which distance it was possible to obtain tightly cropped face images of the required 640×640 resolution.

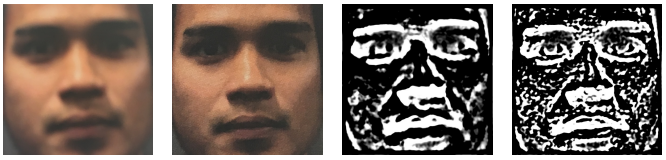

Figure 2. Imposter images captured with an iPhone 6s camera from a distance of 6cm (i) and 9cm (ii). (iii)-(iv) DoG for the images (i)-(ii) with the $\sigma_1 = 4$ and $\sigma_2 = 8$

Imposter images: The imposter images were created from client images as shown in Figure 1 (B). For each subject, an arbitrarily chosen client image, which later would not be used for either training or testing, was printed on A4 paper using a Ricoh 4500 Photocopier. The printed paper was pinned on a board and a series of photos were taken with the camera's autofocus mechanism re-enabled between any two shots.

Before opting for using the autofocus function of the camera we compared between imposter images captured with manual focus and autofocus and found that the latter were more likely to be sharper and thus more challenging to classify correctly. We also note that our settings of the autofocus mechanism prevent the capture of blurry out-of-focus images by blocking the shutter release. Figure 2 shows imposter images produced from small size printed images re-captured by the camera of an iPhone 6s with non-blocking autofocus mechanism. When the autofocus mechanism fails the imposter image becomes extremely blurry, see Figure 2, demonstrating the importance of including the type of camera focus mechanism in the design protocol of the database, for example blocking autofocus, non-blocking autofocus or fixed focus.

Processed imposter images: The processed imposter images were created from the client images as shown in Figure 1 (B). The same arbitrarily chosen client image used to create the imposter images was sharpened using Matlab's *imsharpen* function with standard deviation $\sigma = 8.0$. Then, it was printed on the Ricoh 4500 Photocopier and the same procedure that created the imposter images was followed.

Figure 3 shows instances from the photo-shooting session. The first column of Figure 4 shows client images images from the Durham Face Database. The sharpened client images in the second column, which are not part of the database, exhibit higher contrast and some sharpening artifacts. The third column shows imposter images from the database; they are more blurry than the client images. Finally, the fourth column shows sharpened imposter images from the database, which show the

Figure 3. The photo-shooting sessions. **Left:** Creating client images. **Right:** Creating imposter images.

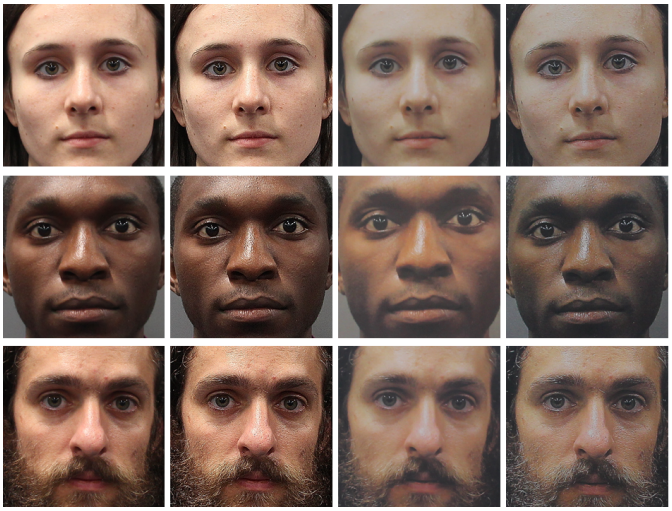

Figure 4. Samples from the Durham Face Database for three different subjects. **Left to right:** a client image; a sharpened client image; an imposter (re-captured) image of the client image; an imposter of the sharpened client.

highest visual similarity with the client images. Indeed, the direct digital sharpening with Matlab's *imsharpen* followed by a procedural blurring by printing them on paper and re-capturing them, seem to a great extent to cancel each other.

4 The liveness test

We tested the database by running on it a well-known liveness test proposed in [2], which is conceptually simple and easy to implement. While testing with various liveness tests would have given us a better understanding of the behaviour of the database, we note that the fundamental nature of the mathematical and statistical tools employed by that test, namely differences of Gaussians of images and sparse logistic regression, make it a suitable choice as a representative liveness test.

Following [2], a difference of Gaussians of the images is used to train a binary classifier with the client class been photos of real faces and the imposter class been recaptured images of the photos of the faces. Assuming that the client and imposter classes are labeled $\{-1, 1\}$, respectively, the conditional probability of an sample x being in the imposter class $y = 1$ is

$$\text{Prob}(y|x) = \frac{1}{1 + \exp(-y(w^T x + b))} \tag{1}$$

where w and b are the weight vector and the intercept returned by the training algorithm. To avoid overfitting, sparse logistic regression is used and thus, the values of w and b are computed through the minimization of the cost function

$$\min_{w,b} \ \text{loss}(w, b) + \lambda \|w\|_1 \tag{2}$$

where the term $\lambda\|w\|_1$ is added to favour sparse weight vectors, i.e. vectors with most of their elements equal to zero, λ is a user defined constant and loss is the standard loss function of the logistic regression

$$\text{loss}(w, b) = \frac{1}{m} \sum_{i=1}^{m} \log(1 + \exp(-y_i(w^T x_i + b))) \tag{3}$$

where m is the size of the training set of samples x_i with associated labels y_i.

5 Testing

We tested the database by running on it the liveness test of Tan et al. [2]. In all tests, the standard deviations of the differences of Gaussians were set at $\sigma_1 = 4$ and $\sigma_2 = 8$. The results are shown in Figure 5. Each diagram consists of two ROC curves, one showing the performance of the classifier in distinguishing between client and imposter images and the other in distinguishing between client and sharpened imposter images. The four diagrams correspond to four different designs of the training set which may contain:

(i) client and imposter images from all 21 subjects,

(ii) client and sharpened imposter images from all 21 subjects,

(iii) client, imposter and sharpened imposter images from all 21 subjects,

(iv) client and imposter images from only 15 of the subjects, while the test set contains images only from the other 6 subjects.

Since client images have always to be included in the training set, Figures 5(i)-(iii) cover all three cases regarding the content of the training set: imposter images, sharpened imposter images, or both. Figure 5(i) verifies our main hypothesis, showing that the performance of the liveness test decreases considerably when the attacker uses sharpened imposter images. Indeed, the large gap between the two curves indicates a significant drop of the performance of the liveness test, which is largely due to the fact that the classifier was trained to distinguish between client and imposter images and not between

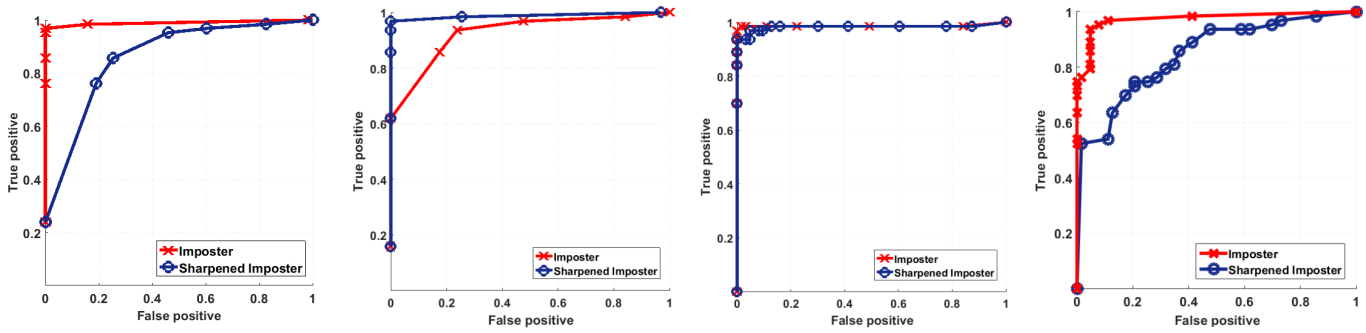

Figure 5. In each diagram the two ROC curves show the performance of the algorithm in distinguishing between client images and imposter or sharpened imposter images, respectively. **(i)-(iii)** The training set consists of client images and: (i) imposter images, (ii) sharpened imposter images, (iii) both imposter and sharpened imposter images, from all 21 subjects. **(iv)** The training set consists of client and imposter images of 15 subjects, while images from the other 6 subjects are used for testing.

client and sharpened imposter images. This can be seen in Figure 5(ii), where the classifier is trained to distinguish between client and sharpened imposter images and as a result the liveness test is much more efficient against attacks with such images. In Figure 5(iii), the classifier is trained to distinguish between client images and imposters of both types and we notice that its discriminative ability is only slightly worse against sharpened imposter images. This is again an expected result since the similarity between sharpened imposter and client images is higher than the similarity between client and imposter images, as it is clear in Figure 4, and thus distinguishing between sharpened imposter and client images is a harder task when no task is given preferential treatment during training.

From the results in Figures 5(i)-(iii) we conclude that by sharpening the client image before printing and re-capturing it to create imposter images, attackers can significantly increase their chances of evading the Tan et al. liveness test. A very simple and largely effective countermeasure is to train the classifier not only with imposter but with sharpened imposter images too. In that case, there is only a slight decrease in the performance of the classifier, which can be explained by the higher similarity between client and sharpened imposter images, which makes the distinction between these two classes an intrinsically more difficult task.

Figure 5(iv) shows the results when the classifier is trained with a cross-subject independent set of client and imposter images. That is, the subjects are partitioned into two non-overlapping subsets and images from the first subset are used for training while images from the second subset are used for testing. We notice that the cross-subject independence is a strong assumption which is not used in the literature since liveness tests usually run in parallel to face recognition systems and a positive classification of a subject by such a system implies the presence of their images in the database. Nevertheless, Figure 5(iv) shows that even with cross-subject independence, our claim that imposter attacks with sharpened images are stronger than common imposter attacks is valid.

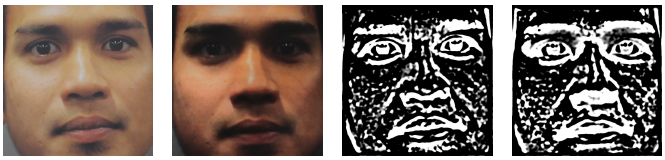

Figure 6. **(i)-(ii)** an imposter images from printouts by a Ricoh 4500 and a Bizhub c654e, respectively. **(iii)-(iv)** the corresponding differences of Gaussians for $\sigma_1 = 4$ and $\sigma_2 = 8$.

5.1 Printer image processing

Apart from pointing out that the strength of malicious imposter attacks can be underestimated if image processing operations before image printing are not considered, the above results are also relevant in the setting of the common imposter attacks through the advanced image processing functionality of modern printers and cameras. As most printers and cameras make use of proprietary image processing technology, the study of the effect of printer and printer settings on the performance of the liveness test is very challenging.

Figure 6 illustrates the effect of printer choice through an example. Indeed, visual inspection of the printouts of the same digital image by two different printers reveals significant differences and as a result the differences of Gaussians of the two images are also significantly different. From each of the two printouts we created 10 imposter images and computed the probabilities of the image to be imposters, see Eq. 1, using the same classifier as in Figure 5(i). When the Ricoh 4500 printout was used, that is, from the printer that created the printouts for the imposter images of the training set, the average probability was 0.929. On the other hand when the Bizhub c654e printer was used the average probability dropped sharply to 0.083. Thus, the example indicates that the choice of printer and printer settings can affect the performance of the liveness test and perhaps, in agreement with what we found in our main experiment, using a variety of printers and printer settings to create the training set can increase the generality of the classifier.

6 Conclusions

In this paper we presented Durham Face Database, a new publicly available face image database for testing anti-spoofing algorithms. The main novelty in its design is the inclusion of processed imposter images, justified under the assumption that attackers may process the client images before printing them for imposter attacks, and in particular, they may sharpen them in order to counteract the blurring induced by the printing and recapturing process. Our tests show a significant decrease in the performance of a standard liveness when sharpened imposter images are used, however, most of the performance loss can be restored by the simple measure of including sharpened imposter images in the training set.

In the future we plan to extend the database with more types of processed imposter images. In particular, instead of processing the client images with existing, simple or sophisticated, image processing operations, we would like to reverse engineer the process and thus, be able to directly compute images which under printing and re-capturing produce imposter images that are as close as possible to the the original client images.

References

[1] L. Omar and I. Ivrissimtzis, "Evaluating the resilience of face recognition systems against malicious attacks," in *BMVW*, 2015.

[2] X. Tan, Y. Li, J. Liu, and L. Jiang, "Face liveness detection from a single image with sparse low rank bilinear discriminative model," in *ECCV*, 2010, pp. 504–517.

[3] L. Sirovich and M. Kirby, "Low-dimensional procedure for the characterization of human faces," vol. 4, pp. 519–524, 1987.

[4] M. Turk and A. Pentland, "Eigenfaces for recognition," vol. 3, no. 1. Cambridge, MA, USA: MIT Press, Jan. 1991, pp. 71–86.

[5] M. A. Turk and A. P. Pentland, "Face recognition using eigenfaces," in *CVPR*. IEEE, 1991, pp. 586–591.

[6] A. Pentland, B. Moghaddam, and T. Starner, "View-based and modular eigenspaces for face recognition," in *CVPR*, Jun 1994, pp. 84–91.

[7] J. Galbally, J. Fierrez, J. Ortega-Garcia, C. McCool, and S. Marcel, "Hill-climbing attack to an eigenface-based face verification system," in *BIdS*, 2009, pp. 1–6.

[8] J. Galbally, C. McCool, J. Fierrez, S. Marcel, and J. Ortega-Garcia, "On the vulnerability of face verification systems to hill-climbing attacks," *Pattern Recognition*, vol. 43, no. 3, pp. 1027–1038, 2010.

[9] S. Marcel, M. S. Nixon, and S. Z. Li, *Handbook of Biometric Anti-Spoofing*. Springer, 2014.

[10] S. Chakraborty and D. Das, "An overview of face liveness detection," *arXiv preprint arXiv:1405.2227*, 2014.

[11] O. Kähm and N. Damer, "2d face liveness detection: An overview," in *BIOSIG*. IEEE, 2012, pp. 1–12.

[12] J. Li, Y. Wang, T. Tan, and A. K. Jain, "Live face detection based on the analysis of fourier spectra," in *Defense and Security*. International Society for Optics and Photonics, 2004, pp. 296–303.

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

[14] K. Kollreider, H. Fronthaler, and J. Bigun, "Non-intrusive liveness detection by face images," *Image and Vision Computing*, vol. 27, no. 3, pp. 233–244, 2009.

[15] C. Lai and C. Tai, "A smart spoofing face detector by display features analysis," *Sensors*, vol. 16, no. 7, p. 1136, 2016.

[16] A. Alotaibi and A. Mahmood, "Enhancing computer vision to detect face spoofing attack utilizing a single frame from a replay video attack using deep learning," in *ICOIP*, 2016, pp. 1–5.

[17] B. Peixoto, C. Michelassi, and A. Rocha, "Face liveness detection under bad illumination conditions," in *IEEE ICIP*, 2011, pp. 3557–3560.

[18] Z. Zhang, J. Yan, S. Liu, Z. Lei, D. Yi, and S. Z. Li, "A face antispoofing database with diverse attacks," in *ICB*. IEEE, 2012, pp. 26–31.

[19] L. Omar and I. Ivrissimtzis, "Resilience of luminance based liveness tests under attacks with processed imposter images," in *WSCG*, 2016.

[20] J. Galbally and R. Satta, "Biometric sensor interoperability: A case study in 3d face recognition."

[21] A. Anjos and S. Marcel, "Counter-measures to photo attacks in face recognition: A public database and a baseline," in *IJCB*, 2011, pp. 1–7.

[22] I. Chingovska, A. Anjos, and S. Marcel, "On the effectiveness of local binary patterns in face anti-spoofing," in *BIOSIG*. IEEE, 2012, pp. 1–7.

[23] J. Maatta, A. Hadid, and M. Pietikainen, "Face spoofing detection from single images using micro-texture analysis," in *IEEE IJCB*, 2011, pp. 1–7.