Spoofing in 2D Face Recognition with 3D Masks and Anti-spoofing with Kinect

Nesli Erdogmus and Sébastien Marcel
Idiap Research Institute
Centre du Parc - rue Marconi 19, CH-1920 Martigny, Suisse
{nesli.erdogmus,sebastien.marcel}@idiap.ch

Abstract

The problem of detecting face spoofing attacks (presentation attacks) has recently gained a well-deserved popularity. Mainly focusing on 2D attacks forged by displaying printed photos or replaying recorded videos on mobile devices, a significant portion of these studies ground their arguments on the flatness of the spoofing material in front of the sensor. In this paper, we inspect the spoofing potential of subject-specific 3D facial masks for 2D face recognition. Additionally, we analyze Local Binary Patterns based countermeasures using both color and depth data, obtained by Kinect. For this purpose, we introduce the 3D Mask Attack Database (3DMAD), the first publicly available 3D spoofing database, recorded with a low-cost depth camera. Extensive experiments on 3DMAD show that easily attainable facial masks can pose a serious threat to 2D face recognition systems and LBP is a powerful weapon to eliminate it.

1. Introduction

With growing populations and their increasing mobility, recognition of humans using biological characteristics becomes a promising solution for identity management. Among many reliable biometric traits, face is a very popular one and it owes this reputation mainly to its accessibility. But unfortunately, this gift can also be a curse in malicious circumstances, enabling attackers to easily create copies and spoof face recognition systems. Spoofing is an attempt to gain authentication through a biometric system by presenting a counterfeit evidence of a valid user [15].

This vulnerability of face has evoked significant attention in the biometric community and numerous papers have been published in countermeasure studies [15]. For a thorough and reproducible analysis of several methods and available public databases, readers can refer to [2, 4].

Due to their convenience and low-cost, the most common types of spoofing methods being focused are photo and video attacks. Proposed anti-spoofing approaches against these attacks can be broadly classified into three groups: liveness detection, motion analysis and texture analysis.

The first group aims to detect liveness of face, based on live-face specific movements such as eye blinking [17] or lip movements [5]. The second group of approaches analyze the motion in the scene and expose spoofing attacks by examining the way the objects move in front of the sensor. The movements of planar objects like papers or screens differ greatly from those of a real face. For this reason, in [10], the trajectories of small regions in face images are analyzed and classified as real or fake. Similarly in [7], a set of facial points are located automatically and their geometric invariants are utilized to detect attacks. Finally, in the third group of methods, the texture of the face image is examined to find spoofing clues like printing artifacts [3] and/or blurring [13]. Alternatively, micro-texture analysis is also applicable as proposed in a recent paper [14] in which multi-scale local binary patterns (LBP) are utilized.

A substantial portion of these approaches for 2D anti-spoofing are rendered inoperative when 3D facial masks are introduced for attacks. For instance in [11], it is shown that a liveness detection system relying on eye-blinking and lip movements can be defeated by simply using photographic masks which are actually high resolution facial prints worn on face with eyes and mouth regions cut out. A similar conclusion is also made in [22]. On the other side, motion-based countermeasures that depend on the shape difference between real and fake faces are not able to operate as intended when the photos or screens are replaced by facial masks. Even employing additional sensors as suggested in the study by Tsalakanidou *et al*. [19], for which a 3D camera is utilized to localize face and test its "face-ness", would become futile in this scenario [8].

In conclusion, it is clear that 3D masks introduce new challenges to face anti-spoofing domain. To our knowledge, very few studies have been published addressing them.

2. Related work

The work of Kim *et al*. [9] can be listed as one of the first papers published in mask anti-spoofing. It aims to distin-

guish between the facial skin and mask materials by exploiting the fact that their reflectance should be different. For this purpose, the distribution of albedo values for illumination at various wavelengths are analyzed to see how different facial skins and mask materials (silicon, latex, or skinjell) behave in reflectance. As a result, a 2D feature vector consisting of 2 radiance measurements under 850 and 685 nm illuminations is selected to be classified via Fisher's linear discriminant. The proposed method is reported to have 97.78% accuracy in fake face detection. In that paper, the experiments are done directly on the mask materials instead of real masks and hence, spoofing performances are not included. Additionally, for mask detection, the measurements are required to be done at exactly 30cm and on the forehead region. The occlusion possibility in the forehead together with range limitations makes the method quite impractical.

Similarly in [22], multi-spectral analysis is proposed claiming that fake, by its definition, is indistinguishable for human eyes and therefore, it is not possible to detect attacks using only visual face images. After measuring the albedo curves of facial skin and mask materials with varying distances, two discriminative wavelengths (850 and 1450 nm) are selected. Finally, an SVM classifier is trained to discriminate between genuine and fake attempts. Experiments are conducted on a database of 20 masks of different materials: 4 plastic, 6 silica gel, 4 paper pulp, 4 plaster and 2 sponge. The results show that the method can achieve 89.18% accuracy. Eliminating the range limitation and experimenting on real facial masks, the authors bring the state of the art one step further, but still no analysis of how well the spoofing attacks work is presented.

These two papers handle the mask attacks in an evasion context rather than spoofing. They don't examine masks that are replicas of real subjects to be impersonated. Contrarily, in [12], Kose *et al*. work on a mask database which consists of printed masks of 16 real subjects. For this purpose, the scans of subjects were acquired by a 3D scanner and the masks were manufactured using a 3D printing service. In addition to texture images, the database also includes range images for both real and fake samples. The authors propose to apply an LBP-based method [14] on both color and depth channels and claim 88.12% and 86% accuracy, respectively. This study has two main shortcomings: Firstly, although they have the means to do so, the authors unfortunately do not report on the spoofing performances of the printed masks. To certify the alleged threat is nearly as important as to counter it. Secondly and more importantly, the utilized database is not public, posing a barrier to comparative and reproducible research.

In our paper, we have two main purposes:

- Introducing a public database, called 3D Mask Attack Database (3DMAD), along with a baseline analysis on its spoofing performance against 2D face recognition
- Studying the effectiveness of Local Binary Patterns (LBP) based features extracted from color and depth images to detect the mask attacks

For reproducibility purposes, both the database[1] and the source code[2] to generate the reported results are made freely available to public use.

The rest of the paper is organized as follows: In Section 3, 3DMAD database is described in detail. In Section 4, the studied countermeasure techniques are explained. Experimental results on 3DMAD for both its capability of deceiving a 2D face recognition system and anti-spoofing performances of the LBP-based methods are provided in Section 5. Finally, in Section 6, the paper is concluded with remarks on future work.

3. The 3D-MAD database

The 3D Mask Attack Database (3DMAD) is mainly composed of real access and mask attack videos of 17 different subjects recorded by Microsoft Kinect sensor. In the following subsections, the database recording is explained in detail and the baseline 2D face recognition algorithm implemented to evaluate the mask spoofing performances is presented.

3.1. 3D mask manufacturing

In [21] Zhang *et al*. state that massive usage of masks does not exist in the literature, mainly due to the fact that it is too expensive to produce client-like masks. This was very true until recently when 3D printing services have sprung up. It has become a market of high potential and is expected to continue growing rapidly.

Among many available options, ThatsMyFace.com stands out with its specialization in facial reconstruction and in transforming 2D portraiture into 3D sculptures. Only after seconds of uploading frontal and profile face images of a person, the constructed 3D face is displayed for inspection. If satisfied, it is printed and delivered to your mailbox in several forms such as a head on an action figure or a wearable life-size mask in hard resin or a paper-cut file.

The advantage of this service over the others is the possibility of utilizing facial images to create a 3D model. Regular 3D printing services like the one used to create the database in [12] require the 3D models to be obtained by the user and uploaded to their system to be printed. Whilst the advancements in 3D scanner technologies are remarkable, they still have range limitations and require user cooperation. For this reason, obtaining proper 3D data from a distance and from unaware subjects is highly unrealistic. On the other hand, photographs of the users can be easily captured form a distance or obtained via Internet, *e.g*. through social networks.

[1] www.idiap.ch/dataset/3dmad

[2] pypi.python.org/pypi/maskattack.lbp

Figure 1. 17 facial masks obtained from ThatsMyFace.com

For our database, we uploaded one frontal and two profile images of 17 different subjects on ThatsMyFace.com and ordered a **life-size wearable mask** and a **paper-cut mask** for each. The uploaded images and the paper-cut masks are also accessible in the database but they are not included in this paper. The 17 wearable masks made out of a hard resin composite in full 24-bit color with holes at the eyes and the nostrils are shown in Figure 1.

The size of the database is limited to 17 subjects due to the high cost of the 3D facial masks. On the other hand, more samples can always be collected from the same masks and additionally, it is possible for everyone to extend it using the available paper-cut files.

3.2. Recording settings

For the dataset, all recordings are done using Microsoft Kinect for Xbox 360. This sensor provides both RGB (8-bit) and depth data (11-bit) of size 640×480 at 30 frames per second. The reason behind this selection is two-fold. Firstly, with the available depth images, it is made possible to explore attacks and devise countermeasures using 3D information. Secondly, it would be interesting to explore the vulnerability of 3D face recognition systems to mask attacks as a future extension to this work.

The videos are collected in three different sessions: Two real-access sessions held two weeks apart and a third session in which mask attacks are performed by a single operator (attacker). In each session and for each person, 5 videos of 10 seconds length are captured. In total, 255 color and depth videos of 300 frames are recorded.

The recording conditions for all three sessions are well-controlled: The background of the scene is uniform and the lighting is adjusted to minimize the shadows cast on the face. Three sample frames from three sessions for the same subject can be seen in Figure 2. Additionally, eye positions for each video are included in the database which are annotated manually at every 60^{th} frame and linearly interpolated for the rest.

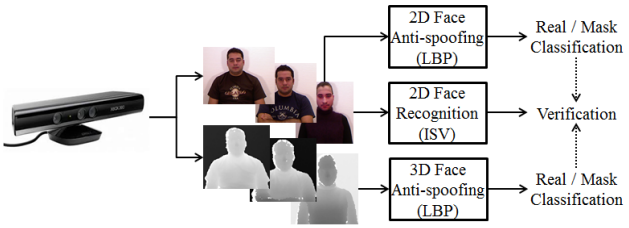

Figure 2. Flowchart for the analyzed methods including sample frames from three different sessions for a subject. The first two are real access samples, while in the third, an attacker is wearing the subject's mask.

3.3. Baseline face recognition algorithm

For completeness, it is necessary to evaluate the spoofing performance of 3D masks on a 2D face recognition system. To this end, Inter Session Variability modeling (ISV) [20] method is implemented as the baseline face recognition algorithm. ISV is an extension of the Gaussian Mixture Models (GMM) approach which estimates more reliable client models by explicitly modeling and removing within-client variations. The identity models are adapted from a Universal Background Model (UBM) and built on Discrete Cosine Transform (DCT) block features.

The recognition tests are done on still images. Specifically, 10 evenly distributed frames are taken from each video. The utilized protocol and obtained results are detailed in Experiments section.

4. LBP-based countermeasures to spoofing

As explained previously, liveness detection and motion analysis methods are bound to fail in detecting 3D mask attacks. This leaves us with one reliable approach which is texture analysis. The fact that human skin differs from mask material with its optical characteristics, such as reflectance, scattering etc. makes it possible to use texture properties to discriminate between real accesses and spoof attacks.

Local Binary Pattern (LBP) [16] and its variations have been proven to be successful in both 2D and 3D face spoofing attacks [6, 12, 14]. In this study, we aim to analyze the discriminative properties of texture features extracted by various LBP operators in 'real face' / '3D mask' classification using the proposed MAD3D database.

Similar to recognition tests, each frame of each video is processed separately to extract the LBP histograms. But differently, the countermeasure analysis is performed per video. To this end, means of all histograms from each video are computed and classification experiments are done on those averaged features.

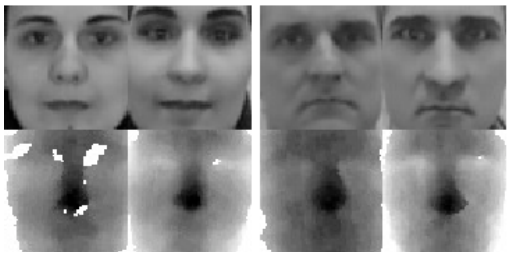

Figure 3. Color (1st row) and depth (2nd row) images after preprocessing. Real access (columns 1 and 3) and mask attack (columns 2 and 4) samples are compared for two subjects.

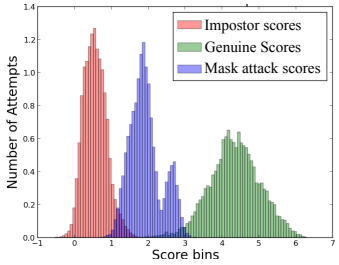

Figure 4. Score distributions of genuine and impostor scores on the development set and mask attack scores on the test set of 3DMAD using ISV for 2D face verification

4.1. Feature extraction by LBP

Basic LBP value for a pixel is computed as a binary number by comparing the intensity of that pixel to the intensities of the adjacent pixels in 3×3 neighborhood ($\text{LBP}_{3\times3}$). The histogram of these 2^8 different labels is then used as a feature vector. In an extension called uniform patterns, the vector length is reduced to 59 by eliminating patterns with more than two bitwise transitions ($\text{LBP}^{u2}_{3\times3}$ - shortly LBP).

In our experiments, along with LBP, three more extensions from [18] are evaluated: transitional (tLBP), direction-coded (dLBP) and modified (mLBP). The tLBP compares two consecutive neighboring pixels circularly in clock-wise direction. The dLBP compares four adjacent pixels only but also includes the direction information in an extra bit. Finally, mLBP compares the pixels in 3×3 neighborhood to their average instead of the center pixel.

Furthermore, we assess the influence of dividing the face images into blocks. For each LBP type, the image is broken into 3×3 blocks, the LBP histograms are calculated for each block separately and concatenated to form the final feature vector. In [14], block processing methodology is reported to improve the performance significantly. On the contrary, it is concluded to be ineffectual in [6].

4.2. Classification

The feature vectors extracted are LBP histograms, so firstly χ^2 histogram matching is applied to compare test samples with a reference histogram which is simply calculated by taking the average of all real access samples in the training set.

Additionally, two more complex classifiers are tested; one being linear and the other non-linear. For linear classification, Linear Discriminant Analysis (LDA) is adopted. Before computing the scores for the extracted features, Principal Component Analysis (PCA) is applied for dimensionality reduction in which 99% of the energy is preserved. Finally for non-linear classification, Support Vector Machine (SVM) with radial kernel basis function is employed.

5. Experiments

Firstly all color and depth frames in the database are preprocessed; i.e. converted to gray-scale, then cropped and normalized to 64×64 (similarly to [6]) using the annotated eye positions. Resulting sample images are given in Figure 3. All experiments are realized using the free signal-processing and machine learning toolbox Bob[3] [1].

5.1. Protocol

The subjects in the database are divided into 3 randomly chosen non-overlapping sets for training, development and testing. Number of identities assigned for each set are 7, 5 and 5, respectively. While experimenting on the database, it is recommended that training (*e.g.* for parameter optimization, building universal models) is done with training and development sets, if needed. The test set should be **solely** used to report performances.

An additional protocol is defined to measure the vulnerability to spoofing for which the development and test sets are further divided into enrollment and probe partitions. While the first sessions are assigned as enrolled gallery samples, the second and third sessions are used for real and mask probing, respectively.

5.2. Spoofing performance of 3D masks

For this experiment, a verification scenario is assumed. After the Universal Background Model is created using the training set, match scores are generated on the probe partitions of development and test sets. The Equal Error Rate (EER) threshold is calculated on the development set as the decision threshold for verification. With this setting, 65.70% of the mask attack attempts in the test set are incorrectly classified as clients. This Spoof False Acceptance Rate (SFAR) validates the mask attacks in 3DMAD as successful spoofing attempts against 2D face recognition.

The false acceptance rate (FAR) at the same threshold would increase from 1.06% to 13.99% if mask attacks are included in the probe partition together with the zero-effort

[3] www.idiap.ch/software/bob

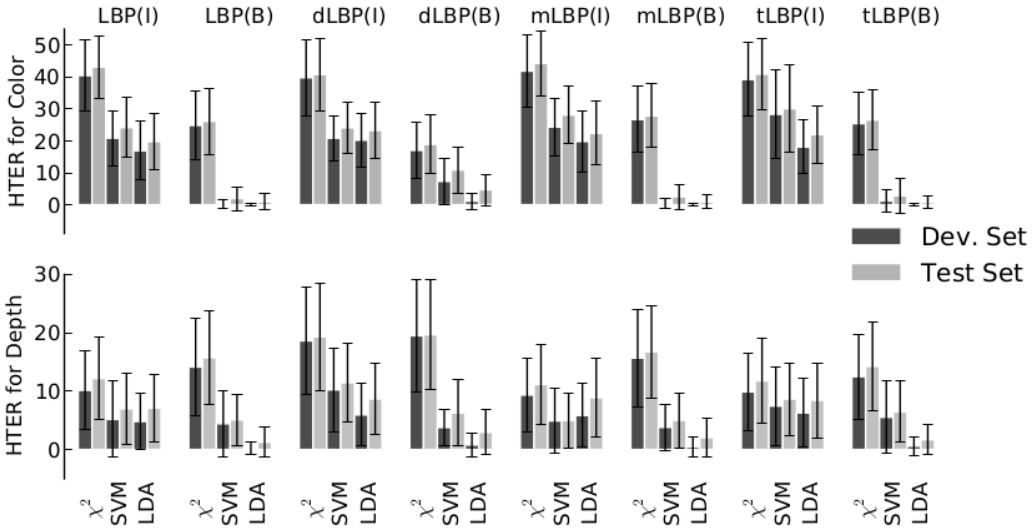

Figure 5. The HTERs of four different LBP types applied per-image (I) and per-block (B) for color and depth images are presented with error bars, where the uncertainties are indicated by standard deviations.

impostor probes. In Figure 4, the score distribution of the mask attacks on the test set is displayed together with the score distributions of genuine clients and impostors on the development set.

The multiple peaks in the mask scores are due to different spoofing capabilities of different masks. This may be affected by the accuracy of the reconstructed 3D model as well as the subject himself.

5.3. Anti-spoofing performances of LBP

Features are extracted for all four LBP types with and without the block-based approach. This results in 8 different sets of features for both color and depth data.

After the classifiers are trained using the training set, the scores are computed for both development and test sets. Since the purpose of the system is binary classification, two types of errors exist: False Fake where the real accesses are classified as mask attacks (FFR) and False Living where the mask attacks are classified as real accesses (FLR). The performances are measured with Half Total Error Rate (HTER) which is the average of this two error rates.

HTERs are calculated at the EER threshold computed on the development set. This threshold is the point along the Receiver Operating Characteristic curve where FFR is equal to FLR.

Due to considerable performance differences among 3D masks, 1000-fold cross validation method is adopted for evaluation. For each fold, the client ids in the database are randomly assigned into one one of the training, development and test sets, respecting their initial sizes. At the end, the error rates are averaged. For reproducibility of the results, a random seed is used to initialize the pseudo-random number generator.

The HTER rates of four different LBP-based methods applied per-image (I) and per-block (B) for both development and test sets are presented in Figure 5 for color and depth images.

If we look at the impact of the block division first, the results show that it improves the results remarkably almost in all settings, except the χ^2 classification for depth maps. For all types of LBP features extracted from range images, per-image approach gives better results when χ^2 is used.

For the classification techniques, in general LDA yields to better performances for both color and depth features. The exceptions occur when LBP and mLBP features are extracted from depth images without using blocks. In those cases, SVM marginally draws ahead of LDA.

Finally, the experiments reveal that it is not easy to pick one method among four LBP types, giving best results for all different settings. For the color images, LBP performs better than the rest with LDA and SVM classifiers, whereas with χ^2 dLBP overcomes. For the depth images, with χ^2 and LDA, tLBP gives smaller errors in average. On the other hand for SVM, mLBP is better.

Results on this database suggest a general trend that classification of block-based LBP features with LDA gives best results with both color and depth images, for which HTER values are found to be 0.95% and 1.27%, respectively.

6. Conclusion

Utilization of 3D masks in spoofing attacks becomes easier / cheaper each day with the advancements in 3D printing technology. In this paper, we aim to contribute to the current state of research in this domain; by presenting a novel pub-

lic database of 3D mask attacks accompanied by protocols and a baseline 2D face recognition system that is proved to be vulnerable to those attacks, and by giving an analysis on various LBP-based anti-spoofing methods using color and depth images obtained from Kinect. The experimental results generally suggest that for both data types, LDA classification of block-based extracted uniform LBP features is more accurate in mask detection.

For the sake of reproducible research, the source code is made publicly available, together with the database and its protocols. A possible extension to this work is to explore the spoofing performances in 3D face recognition systems and to devise methods to detect attacks using pure 3D data, instead of 2.5D. Additionally, further investigation can be done on spoofing the spoofing potential of each mask separately.

Acknowledgments

The authors would like to express their thanks to the FP7 European TABULA RASA Project (257289) for the financial support and to the partners within the consortium for their fruitful collaboration.

References

[1] A. Anjos, L. El-Shafey, R. Wallace, M. Günther, C. McCool, and S. Marcel. Bob: a free signal processing and machine learning toolbox for researchers. In *ACM International Conference on Multimedia*, pages 1449–1452, 2012.

[2] A. Anjos and S. Marcel. Counter-measures to photo attacks in face recognition: A public database and a baseline. In *International Joint Conference on Biometrics*, pages 1 –7, October 2011.

[3] J. Bai, T.-T. Ng, X. Gao, and Y.-Q. Shi. Is physics-based liveness detection truly possible with a single image? In *IEEE International Symposium on Circuits and Systems*, pages 3425 –3428, June 2010.

[4] M. Chakka, A. Anjos, S. Marcel, R. Tronci, D. Muntoni, G. Fadda, M. Pili, N. Sirena, G. Murgia, M. Ristori, F. Roli, J. Yan, D. Yi, Z. Lei, Z. Zhang, S. Li, W. Schwartz, A. Rocha, H. Pedrini, J. Lorenzo-Navarro, M. Castrillon-Santana, J. Maatta, A. Hadid, and M. Pietikainen. Competition on counter measures to 2-d facial spoofing attacks. In *International Joint Conference on Biometrics*, pages 1 –6, October 2011.

[5] G. Chetty and M. Wagner. Multi-level liveness verification for face-voice biometric authentication. In *Biometrics Symposium: Special Session on Research at the Biometric Consortium Conference*, pages 1–6, 2006.

[6] I. Chingovska, A. Anjos, and S. Marcel. On the effectiveness of local binary patterns in face anti-spoofing. In *International Conference of Biometrics Special Interest Group*, pages 1 –7, September 2012.

[7] M. De Marsico, M. Nappi, D. Riccio, and J. Dugelay. Moving face spoofing detection via 3d projective invariants. In *IAPR International Conference on Biometrics*, pages 73 –78, April 2012.

[8] N. Erdogmus and S. Marcel. Spoofing 2d face recognition systems with 3d masks. In *Conference of the Biometrics Special Interest Group (BIOSIG)*, 2013.

[9] Y. Kim, J. Na, S. Yoon, and J. Yi. Masked fake face detection using radiance measurements. *Journal of the Optical Society of America A*, 26(4):760–766, 2009.

[10] K. Kollreider, H. Fronthaler, and J. Bigun. Evaluating liveness by face images and the structure tensor. In *IEEE Workshop on Automatic Identification Advanced Technologies*, pages 75 – 80, October 2005.

[11] K. Kollreider, H. Fronthaler, and J. Bigun. Verifying liveness by multiple experts in face biometrics. In *IEEE Computer Society Conference on Computer Vision and Pattern Recognition Workshops*, pages 1 –6, June 2008.

[12] N. Kose and J.-L. Dugelay. Countermeasure for the protection of face recognition systems against mask attacks. In *IEEE International Conference on Automatic Face and Gesture Recognition*, April 2013.

[13] J. Li, Y. Wang, T. Tan, and A. K. Jain. Live face detection based on the analysis of fourier spectra. In *SPIE 5404, Biometric Technology for Human Identification*, pages 296–303, 2004.

[14] J. Maatta, A. Hadid, and M. Pietikainen. Face spoofing detection from single images using micro-texture analysis. In *International Joint Conference on Biometrics*, pages 1–7, 2011.

[15] K. Nixon, V. Aimale, and R. Rowe. Spoof detection schemes. In A. Jain, P. Flynn, and A. Ross, editors, *Handbook of Biometrics*, pages 403–423. Springer US, 2008.

[16] P. M. Ojala T. and M. T. Multiresolution gray-scale and rotation invariant texture classification with local binary patterns. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 24(7):971–987, 2002.

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

[18] J. Trefnỳ and J. Matas. Extended set of local binary patterns for rapid object detection. In *Proceedings of the Computer Vision Winter Workshop*, 2010.

[19] F. Tsalakanidou, C. Dimitriadis, and S. Malassiotis. A secure and privacy friendly 2d+3d face authentication system robust under pose and illumation variation. In *International Workshop on Image Analysis for Multimedia Interactive Services*, page 40, June 2007.

[20] M. C. Wallace R., McLaren M. and M. S. Inter-session variability modelling and joint factor analysis for face authentication. In *International Joint Conference on Biometrics*, pages 1–8, 2011.

[21] Z. Zhang, J. Yan, S. Liu, Z. Lei, D. Yi, and S. Li. A face antispoofing database with diverse attacks. In *IAPR International Conference on Biometrics*, pages 26–31, 2012.

[22] Z. Zhang, D. Yi, Z. Lei, and S. Li. Face liveness detection by learning multispectral reflectance distributions. In *IEEE International Conference on Automatic Face Gesture Recognition and Workshops*, pages 436 –441, March 2011.