# Face Anti-Spoofing using Haralick Features

Akshay Agarwal, Richa Singh, and Mayank Vatsa
IIIT-Delhi, India
{akshaya, rsingh, mayank} @iiitd.ac.in

## Abstract

*Face spoofing can be performed in a variety of ways such as replay attack, print attack, and mask attack to deceive an automated recognition algorithm. To mitigate the effect of spoofing attempts, face anti-spoofing approaches aim to distinguish between genuine samples and spoofed samples. The focus of this paper is to detect spoofing attempts via Haralick texture features. The proposed algorithm extracts block-wise Haralick texture features from redundant discrete wavelet transformed frames obtained from a video. Dimensionality of the feature vector is reduced using principal component analysis and two class classification is performed using support vector machine. Results on the 3DMAD database show that the proposed algorithm achieves state-of-the-art results for both frame-based and video-based approaches, including 100% accuracy on video-based spoofing detection. Further, the results are reported on existing benchmark databases on which the proposed feature extraction framework archives state-of-the-art performance.*

## 1. Introduction

Attacks on a face recognition system can have adverse effects including impersonating someone or eluding one's own identity, particularly using disguise [10] and spoofing attacks [21]. It has been established that spoofing attacks pose vulnerability to face recognition systems [2,6,7,11–13, 19,23,24] as the cost of such spoofing attempts are not very high. Among face spoofing attempts, traditionally, print and replay attacks are prevalent; however, with advancement in technology (e.g. 3D printing) and their cost effectiveness, printing a 3D face mask is easy. These 3D mask, made with care, are *real-like* and can be comfortably used for deceiving face recognition systems. Figure 1 shows examples of print, replay, and 3D masks attacks from three popular face spoofing databases, CASIA-FASD [29], 3DMAD [12], and MSU-MFSD [27]. It is imperative that face recognition systems should be equipped with countermeasures, particularly for law enforcement and for moderate to important security applications.

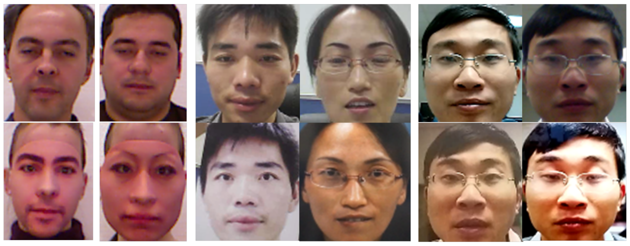

Figure 1: Samples of spoofed and non-spoofed images from CASIA-FASD [29], 3DMAD [12], and MSU-MFSD [27] databases. First two column samples are from 3DMAD, next two column samples are from CASIA-FASD, and last two column samples are from MSU-MFSD databases.

In literature, different anti-spoofing approaches have been proposed and a summary of some of the recent approaches are presented in Table 1. This table shows that approaches ranging from simple image quality measures to local binary pattern (LBP) and other texture descriptors based approaches to deep learning architectures are explored. Inspired from *Occam's Razor*, in this paper we present a *simple-yet-effective* face anti-spoofing algorithm which extracts Haralick features [17] from redundant discrete wavelet decomposed video frames, followed by principal component analysis based dimensionality reduction and SVM classification. Experiments on the 3DMAD database [12], CASIA-FASD [29], and MSU-MFSD [27] show state of the art performance on video face anti-spoofing with very high computational efficiency, both during training and testing.

## 2. Proposed Algorithm

The proposed anti-spoofing algorithm is based on texture features extracted using redundant discrete wavelet transform (RDWT) and Haralick descriptor. While Haralick is a robust texture feature descriptor, RDWT provides the multi-scale decomposition with over-complete representation. Hua and Fowler [18] have argued that unlike the DWT, RDWT is shift invariant, its redundancy introduces

Table 1: Summarizing previous research in face anti-spoofing.

| **Authors** | **Database** | **Contribution** |
|---|---|---|
| Erdogmus and Marcel [13] (2014) | Morpho and 3DMAD | LBP, tLBP, mLBP, dLBP + SVM (linear and RBF), $\chi^2$, and LDA |
| Yi *et al.* [28] (2014) | In-house | Multi-Spectral+ VIS and NIR |
| Pereira *et al.* [9] (2014) | CASIA, Replay Attack | LBP-TOP + LDA and SVM (RBF) |
| Galbally *et al.* [15] (2014) | Multiple databases | Full reference and no reference image quality assessment + LDA and QDA |
| Anjos *et al.* [1] (2014) | In-house | Face spoofing modalities are presented and different counter measures in visual spectrum are analyzed |
| Galbaly and Marcel [14] (2014) | Replay attack and CASIA | Image Quality + LDA |
| Menotti *et al.* [22] (2015) | Multiple databases | Deep CNN + hard margin linear SVM (architecture and filter optimization) |
| Wen *et al.* [27] (2015) | In-house | Distortion analysis feature vector (specular reflection, color distribution ) + Ensemble of classifier + frame fusion |
| Hadid *et al.* [16] (2015) | – | Various spoofing attacks on the biometric systems and the existing anti-spoofing measures are reviewed |
| Patel *et al.* [25] (2015) | Replay Attack, CASIA and In-house(RAFS) | Moire Pattern detection based on multi-scale LBP and DSIFT + SVM (RBF) |

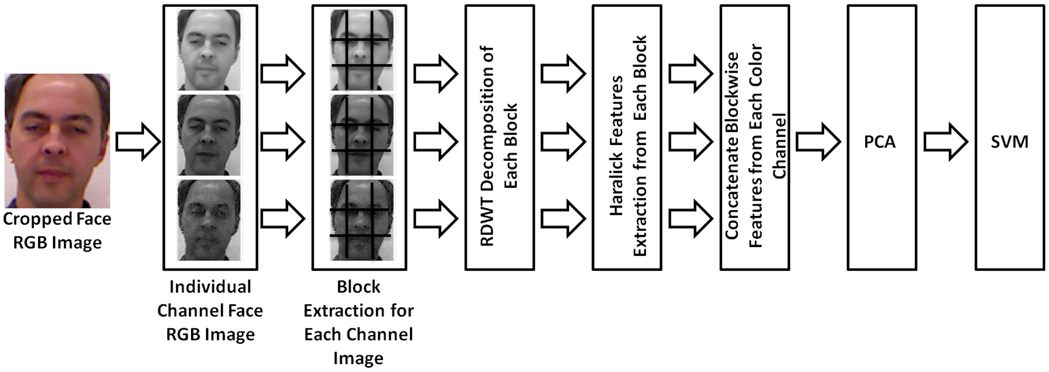
Cropped Face RGB Image
Individual Channel Face RGB Image
Block Extraction for Each Channel Image
RDWT Decomposition of Each Block
Haralick Features Extraction from Each Block
Concatenate Blockwise Features from Each Color Channel
PCA
SVM

Figure 2: Illustrating the steps involved in the proposed algorithm for face spoofing detection.

an over-complete frame expansion, and functions as a better approximation to the continuous wavelet transform. It is our observation that the minute differences in spoofed and non-spoofed videos are accentuated when Haralick features are extracted from RDWT sub-bands. Figure 2 shows the steps involved in the proposed spoofing detection algorithm. Since the entire image contains background along with face images, the first step in the pipeline is detecting facial regions from the frames. In the proposed pipeline, no other preprocessing has been applied, except normalizing frames to a fixed size using bicubic interpolation.

At the time of acquisition, if spoofed properly using 3D mask, print or replay attack, there should not be visible variations in the real and spoofed faces, except micro level variations. However, based on the reflectance properties and texture of the spoof material, it is our assertion that we can accentuate the differences in the wavelet domain, with the correct choice of the mother wavelet. Further, due to the variations in capture device, camera focus, and surrounding illumination, there can be differences at local regions in the image. Therefore, the color image is first decomposed into individual RGB channels and each channel image is divided into non-overlapping local patches of size $32 \times 32$. Every local patch is then transformed into wavelet domain using single level redundant discrete wavelet transform. Once the decomposed RDWT sub-bands are obtained, the next step is to extract the features. In this research, we extract Haralick features [17] which measures the pixel texture distribution within the local neighborhood and encodes the variations in pixel intensities and the homogeneity of the image. Haralick features are extracted over each patch individually. The steps involved in feature extraction are listed below:

1. Image is divided into non-overlapping patches,

2. Over each patch, RDWT decomposition is performed to obtain four sub-bands: approximation ($H_a$), hori-

zontal ($H_h$), vertical ($H_v$), and diagonal ($H_d$),

3. Each sub-band in RDWT is of the same size as the original patch. Haralick features are computed over all the four sub-bands for all patches,

4. Haralick features [17] are also computed over the original image patch without wavelet decomposition,

5. Final feature vector for one color channel is the concatenation of Haralick features over 4 sub-bands of each patch and all the original patch.

Steps 1 to 5 are repeated for each color channel separately and the final feature vector is obtained by concatenating all the features computed over red, green and blue color channels. The 13 Haralick features computed in this research are listed below in Table 2.

Table 2: Haralick features used for face anti-spoofing.

| **Haralick Features** | |
|---|---|
| Angular Second Moment | Contrast |
| Correlation | Sum of Squares: Variance |
| Inverse Difference Moment | Sum Average |
| Sum Variance | Sum Entropy |
| Entropy | Difference Variance |
| Difference Entropy | Info. Measure of Correlation 1 |
| Info. Measure of Correlation 2 | – |

The feature extraction algorithm described above yields a feature vector of size 2340 for one image/frame. For frame classification, such feature dimensionality is computationally feasible. However, for video classification, feature vectors of size 2340 are computed from each frame and then concatenated. Since videos in spoofing detection problem generally contain $\sim$300 frames, i.e. 10 seconds with 30 fps. This leads to a feature vector of large dimensionality and learning a classifier with such high dimensional feature vector requires a lot of training data points. Therefore, for video-based face anti-spoofing, dimensionality reduction is performed using principal component analysis (PCA) to preserve principal components corresponding to 99% Eigenenergy. With the final feature vector, a two-class support vector machine classifier [26] is learnt to classify the features into *spoof* or *non-spoof* classes.

## 3. Experimental Protocol and Results

The experiments are performed on three popular face spoofing databases: 3DMAD [12], CASIA-FASD [29], and MSU-MFSD [27].

### 3.1. Experiments on 3DMAD Database

Recently, two different efforts have shown that use of 3D mask is a viable way of spoofing face recognition algorithms [13, 20]. Therefore, the efficacy of the proposed algorithm is demonstrated on the 3DMAD [12] face spoofing database that contains spoofed and non-spoofed videos corresponding to 17 subjects. The database is collected using Kinect sensor in three different sessions and 5 videos of each subject are collected in each session. Two session videos correspond to real access and the third session videos are mask attack videos. Thus, the database contains a total of 10 real face videos of each person and 5 videos of each subjects using the 3D mask attack. As per the benchmark protocol, the database is divided into three sets: 1) training set, 2) development set, and 3) testing set. 17 fold leave one out cross validation (LOOCV) is performed. It is ensured that the subjects used in each of the three sets are mutually exclusive. For each fold, data belonging to one person goes to the testing set and the data of remaining subjects is divided into training and validation, 50% each. The SVM classifier (LibSVM [5] is utilized) is optimized using grid search in terms of the parameters and kernels. The grid search shows that linear kernel yields the best results on both training and validation datasets.

Frames from each of the videos are obtained and face region is extracted using the eye-coordinates provided with the database. The features explained above are extracted from the RGB frames. As per the protocol, we demonstrate the results on both frame classification and video classification. For frame classification, each frame of the video is classified either as *non-spoof* or *spoof*. The results are reported in terms of both correct classification accuracy and Half Total Error Rate (HTER). HTER is computed as the average of False Living Rate and False Fake Rate. Table 3 and Figure 3 summarizes the results of frame classification using the proposed algorithm without block classification and with block-wise classification. The proposed algorithm involves concatenating features from individual R, G, and B channels. Therefore, we have also evaluated the performance of individual channels, gray scale image, and depth maps (given as part of the data) as well.

On the testing set, the proposed algorithm, without dividing into blocks, yields 4.1% HTER. With block based version of the proposed algorithm yields 3% HTER. It is interesting to note that only using Red channel with blockwise feature extraction yields 0% HTER on both validation and testing sets. We believe that this is because, among the three channels, R channel is close to NIR spectrum (red = 680nm, NIR = 700 - 900nm) and therefore, shows close absorption properties. When a different material is used for spoofing, R channel encodes these variations better than other two channels hence, better results are obtained.

The results of video based anti-spoofing are reported with and without applying PCA and block-wise feature extraction. Table 4 summarizes the results of the proposed algorithm.It can be observed that dimensionality reduction not only helps in reducing the feature dimension but also

Table 3: Frame based face anti-spoofing results with Haralick features on 3DMAD database. Results are reported in terms of classification accuracy (%) and half total error rate (%).

| Algorithm | Image Channel | Classification Accuracy | | HTER | |
|---|---|---|---|---|---|
| | | Validation | Testing | Validation | Testing |
| Without block-wise | **RGB** | 90.2 | **95.8** | 10.1 | **4.1** |
| | **R** | **90.9** | 94.4 | **7.1** | 4.3 |
| | G | 77.7 | 82.5 | 27.9 | 20.0 |
| | B | 82.5 | 86.3 | 18.4 | 14.6 |
| | GrayScale | 79.3 | 79.8 | 24.7 | 22.1 |
| | Depth | 79.2 | 81.6 | 23.9 | 20.2 |
| With block-wise | RGB | 95.9 | 97.8 | 6.0 | 3.0 |
| | **R** | **100** | **100** | **0.0** | **0.0** |
| | G | 96.0 | 98.0 | 5.0 | 2.0 |
| | B | 96.0 | 97.6 | 5.0 | 3.0 |

Table 4: Video based face anti-spoofing results with Haralick features on 3DMAD database. Results are reported in terms of classification accuracy (%) and half total error rate (%).

| Algorithm | Image Channel | Classification Accuracy | | HTER | |
|---|---|---|---|---|---|
| | | Validation | Testing | Validation | Testing |
| Without block-wise + PCA | **RGB** | **97.5** | **97.6** | **2.0** | **2.0** |
| | R | 79.2 | 85.9 | 16.0 | 10.0 |
| | G | 68.9 | 77.2 | 41.0 | 28.0 |
| | B | 78.0 | 83.9 | 25 | 18.0 |
| | Gray scale | 70.0 | 80.3 | 34.0 | 21.0 |
| | Depth image | 71.9 | 74.5 | 33.0 | 29.0 |
| With block-wise + PCA | **RGB** | **100** | **100** | **0.0** | **0.0** |
| | R | 99.9 | **100** | **0.0** | **0.0** |
| | G | 96.1 | 98.0 | 5.0 | 2.0 |
| | B | 98.8 | 99.2 | 1.0 | **0.0** |

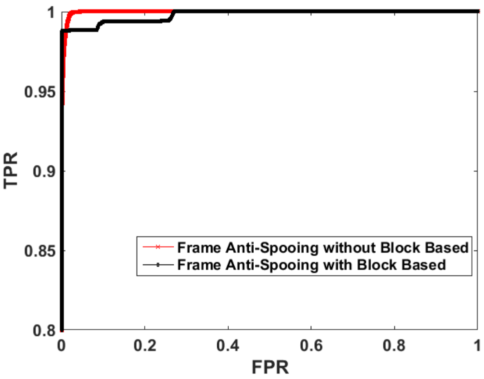

Figure 3: ROC curves for frame-wise anti-spoofing on the 3DMAD database.

improves the accuracy of video anti-spoofing. As shown in the Table 4, 0% HTER and 100% accuracy is obtained with video based anti-spoofing.

Each video in 3DMAD database has 300 frames and therefore, the size of feature vector is same. However, if the length of the spoof and real video is different then the concatenation of the features vector of all frames can yield different feature dimension. To address this problem, in place of concatenating features from all frames into one final feature vector, individual frames are classified and an average score of all frames is computed as the final score of the video. This score level fusion approach yields the EER and HTER of 0% and 0.59%, respectively.

If the video length is longer, then the computation of the features and classification of the video using all frames can be time-consuming. We perform additional experiments to evaluate how many frames are necessary enough to detect a spoofing attempt. 30 frames are selected from each video and the proposed feature extraction is applied. Each selected frame is classified using SVM and the score is calculated. The final score of a video is calculated by taking the average of all 30 frames scores. The results in Table

5 show that on 3DMAD database, 30 frames are sufficient for anti-spoofing. Further, to mitigate 3D mask attack, local binary pattern (LBP) based texture features along with support vector machine (SVM) based classification [11] has shown very high accuracies. Recently, Menotti *et al.* [22] propose to use deep learning architecture to achieve near-perfect accuracy on the 3DMAD database [12]. Table 5 shows the comparison of the proposed algorithm with these two existing algorithms which illustrates that the proposed algorithm is as efficient as them.

In terms of time, on 3DMAD database, feature extraction process requires 150 milliseconds per frame. Once the features are extracted, video-wise classification requires 400 milliseconds with PCA and 620 milliseconds without PCA on a 3.4 GHz $i7$ desktop with 16GB RAM in Matlab programming environment. Computationally, the proposed algorithm is not very resource-hungry and can be trained on a desktop machine.

Table 5: Comparison between the proposed and state-of-the-art methods on the 3DMAD database.

| **Algorithm** | **# Frames** | **EER %** | **HTER %** |
|---|---|---|---|
| Erdogmus and Marcel [13] (2014) | All | 0 | 0 |
| Menotti *et al.* [22] (2015) | All | – | 0 |
| Proposed | 30 | 0 | **0.29** |
| | All | 0 | **0** |

### 3.2. Experiments on CASIA-FASD Database

To show the effectiveness of the proposed methodology, we have performed experiments with CASIA face spoofing database as well. The database consists of videos pertaining to 50 subjects in which 20 subjects are used for training and 30 subjects are used for testing as defined in the protocol. Total 12 videos are recorded for each subject with 3 different camera resolution: low, normal and high. Different kinds of attacks are performed while performing the spoofing attack. The attacks performed are: Warped photo, CutPhoto, and Video attack. There are seven testing protocols [29]: (1) low-resolution test, (2) normal resolution test, (3) high-resolution test, (4) warped photo attack test, (5) cut-photo attack test, (6) video attack test, and (7) overall test. Overall results combined all the quality and attack sets to report the overall performance of the system. Similar to 3DMAD experiments, every face image is divided into blocks and the proposed features are extracted. SVM model is trained using the training dataset and for testing, individual frames are classified as genuine or spoof and a score is computed. Since, in CASIA-FASD database, number of frames in each video is different, the final score is computed by taking the average of all frame scores. The proposed algorithm yields an equal error rate (EER) of 6.7% with 30 frames experiment and 1.1% with all frames experiment. As shown in Table 6, db1 mother wavelet in RDWT yields better result compared to bior5.5 and RDWT decomposition achieves better accuracy than DWT (these are the observations on 3DMAD as well). Further, Table 7 shows that current best-reported EER on the CASIA-FASD database is 3.2% [4] and the proposed algorithm achieves state of the art results with 1.1% EER. We further analyze the performance with respect seven protocols in CASIA-FASD. The proposed algorithm achieves 0% EER on low and normal quality test; however, the most challenging attack is cut-photo attack followed by video attack.

Table 6: Study of different wavelet decomposition and wavelet filters on the CASIA-FASD database using “30 frames experiment”.

| **Wavelet Decomposition** | **Wavelet Filter** | **EER %** |
|---|---|---|
| DWT | db1 | 9.81 |
| | bior5.5 | 8.88 |
| RDWT | **db1** | **6.66** |
| | bior5.5 | 10.00 |

Table 7: Comparison between the proposed and state-of-the-art methods on the CASIA-FASD database.

| **Algorithm** | **# Frames** | **EER %** |
|---|---|---|
| Pereira et al. [8] (2012) | All | 10.6 |
| Bharadwaj et al. [3] (2014) | All | 14.4 |
| Boulkenafet et al. [4] (2016) | All | 3.2 |
| Proposed | 30 | 6.7 |
| | All | **1.1** |

### 3.3. Experiments on MSU-MFSD Database

The performance of the proposed algorithm is also evaluated on the MSU mobile face spoof database. Using the pre-defined experimental protocol [27], results of the proposed algorithm and comparison with existing approaches are reported in the Table 8. Similar to other results, the proposed algorithm enhances the state of art results on this database as well.

Table 8: Comparison between the proposed and state-of-the-art method on the MSU-MFSD database.

| **Algorithm** | **# Frames** | **EER %** |
|---|---|---|
| Wen et al. [27] (2015) | All | 8.58 |
| Boulkenafet et al. [4] (2016) | All | 3.5 |
| Proposed | 30 | **2.9** |
| | All | 5.0 |

## 4. Conclusion

3D masks, print attacks, and replay attacks have been shown to spoof face recognition algorithms and therefore, it is imperative to have a *pre-processing* step to detect if input video/frame is spoofed or not. In this paper, we present a novel, cost-effective and simple anti-spoofing algorithm which extracts block-wise Haralick features from RDWT sub-bands. For video anti-spoofing, features obtained from each frame are concatenated and dimensionality is reduced using PCA followed by SVM classification. Experiments on 3DMAD, CASIA-FASD, and MSU-MFSD face spoofing databases show that the proposed algorithm achieves state-of-the-art performance for video based anti-spoofing. In future, we plan to showcase the results of cross-database experiments.

## 5. Acknowledgement

This research is supported by Vishveshwarya PhD Fellowship, Government of India.

## References

[1] A. Anjos, J. Komulainen, S. Marcel, A. Hadid, and M. Pietikäinen. Face anti-spoofing: Visual approach. In *Handbook of Biometric Anti-Spoofing*, pages 65–82. Springer, 2014.

[2] A. Anjos and S. Marcel. Counter-measures to photo attacks in face recognition: A public database and a baseline. In *IJCB*, 2011.

[3] S. Bharadwaj, T. I. Dhamecha, M. Vatsa, and R. Singh. Face anti-spoong via motion magnication and multifeature videolet aggregation. Technical report, 2014. Available at `https://repository.iiitd.edu.in/jspui/handle/123456789/138`.

[4] Z. Boulkenafet, J. Komulainen, and A. Hadid. Face spoofing detection using colour texture analysis. *IEEE TIFS*, PP(99):1–1, 2016.

[5] C.-C. Chang and C.-J. Lin. LIBSVM: A library for support vector machines. *ACM TIST*, 2:27:1–27:27, 2011. Software available at `http://www.csie.ntu.edu.tw/~cjlin/libsvm`.

[6] G. Chetty and M. Wagner. Multi-level liveness verification for face-voice biometric authentication. In *Biometric Consortium Conference*, pages 1–6, Sept 2006.

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

[8] T. de Freitas Pereira, A. Anjos, J. M. De Martino, and S. Marcel. Lbp- top based countermeasure against face spoofing attacks. In *ACCV 2012 Workshops*, pages 121–132. Springer, 2012.

[9] T. de Freitas Pereira, J. Komulainen, A. Anjos, J. M. De Martino, A. Hadid, M. Pietikainen, and S. Marcel. Face liveness detection using dynamic texture. *EURASIP JIVP*, 2014:2, Jan. 2014.

[10] T. Dhamecha, A. Nigam, R. Singh, and M. Vatsa. Disguise detection and face recognition in visible and thermal spectrums. In *ICB*, pages 1–8, June 2013.

[11] N. Erdogmus and S. Marcel. Spoofing 2d face recognition systems with 3d masks. In *BIOSIG*, pages 209–216, 2013.

[12] N. Erdogmus and S. Marcel. Spoofing in 2d face recognition with 3d masks and anti-spoofing with kinect. In *IEEE BTAS*, pages 1–6, Sept 2013.

[13] N. Erdogmus and S. Marcel. Spoofing face recognition with 3d masks. *IEEE TIFS*, 9(7):1084–1097, July 2014.

[14] J. Galbally and S. Marcel. Face anti-spoofing based on general image quality assessment. In *22nd ICPR*, pages 1173–1178, Aug 2014.

[15] J. Galbally, S. Marcel, and J. Fierrez. Image quality assessment for fake biometric detection: Application to iris, fingerprint, and face recognition. *IEEE TIP*, 23(2):710–724, 2014.

[16] A. Hadid, N. Evans, S. Marcel, and J. Fierrez. Biometrics systems under spoofing attack: An evaluation methodology and lessons learned. *IEEE SPL*, 32(5):20–30, Sept 2015.

[17] R. Haralick, K. Shanmugam, and I. Dinstein. Textural features for image classification. *IEEE TSMC*, 3(6):610–621, Nov 1973.

[18] L. Hua and J. E. Fowler. Rdwt and image watermarking. Technical report, 2001.

[19] 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, Oct 2005.

[20] N. Kose and J.-L. Dugelay. Countermeasure for the protection of face recognition systems against mask attacks. In *IEEE FG*, pages 1–6, April 2013.

[21] S. Marcel, M. S. Nixon, and S. Z. Li. *Handbook of Biometric Anti-Spoofing: Trusted Biometrics Under Spoofing Attacks*. Springer Publishing Company, 2014.

[22] D. Menotti, G. Chiachia, A. Pinto, W. Robson Schwartz, H. Pedrini, A. Xavier Falcao, and A. Rocha. Deep representations for iris, face, and fingerprint spoofing detection. *IEEE TIFS*, 10(4):864–879, April 2015.

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

[24] G. Pan, L. Sun, Z. Wu, and S. Lao. Eyeblink-based anti-spoofing in face recognition from a generic webcamera. In *IEEE ICCV*, pages 1–8, Oct 2007.

[25] K. Patel, H. Han, A. Jain, and G. Ott. Live face video vs. spoof face video: Use of moire patterns to detect replay video attacks. In *ICB*, pages 98–105, May 2015.

[26] V. N. Vapnik. *The Nature of Statistical Learning Theory*. Springer-Verlag New York, Inc., 1995.

[27] D. Wen, H. Han, and A. Jain. Face Spoof Detection with Image Distortion Analysis. *IEEE TIFS*, 10(4):746–761, April 2015.

[28] D. Yi, Z. Lei, Z. Zhang, and S. Z. Li. Face anti-spoofing: Multi-spectral approach. In *Handbook of Biometric Anti-Spoofing*, pages 83–102. Springer, 2014.

[29] Z. Zhang, J. Yan, S. Liu, Z. Lei, D. Yi, and S. Li. A face antispoofing database with diverse attacks. In *ICB*, pages 26–31, March 2012.