Scale space texture analysis for face anti-spoofing

Zinelabidine Boulkenafet[1], Jukka Komulainen[1], Xiaoyi Feng[2], Abdenour Hadid[1,2]
[1] Center for Machine Vision Research, University of Oulu, Finland
[2] Northwestern Polytechnical University, School of Electronics and Information, Xian, China

Abstract

Face spoofing detection (i.e. face anti-spoofing) is emerging as a new research area and has already attracted a good number of works during the past five years. This paper addresses for the first time the key problem of the variation in the input image quality and resolution in face anti-spoofing. In contrast to most existing works aiming at extracting multiscale descriptors from the original face images, we derive a new multiscale space to represent the face images before texture feature extraction. The new multiscale space representation is derived through multiscale filtering. Three multiscale filtering methods are considered including Gaussian scale space, Difference of Gaussian scale space and Multiscale Retinex. Extensive experiments on three challenging and publicly available face anti-spoofing databases demonstrate the effectiveness of our proposed multiscale space representation in improving the performance of face spoofing detection based on gray-scale and color texture descriptors.

1. Introduction

There has been a significant progress in face recognition research during the last two decades [18]. For instance, the performance of some methods on the well-known Labeled Faces in the World (LFW) database surpasses the human performance [11]. As a result, face recognition systems started to be deployed in real-world applications such as mobile device authentication, identity card management, and security portal verification. As a matter of fact, these face recognition systems rely on data which are personal in nature but, nevertheless, are already public (e.g. social media) or can be easily stolen (e.g. captured). These public or stolen data can be used to gain an illegitimate access to the systems by presenting them, on printed paper or replayed on display devices, in front of the camera, which is known as spoofing attacks or presenting attacks. Inspecting the resilience of several commercial face recognition systems against spoofing attacks [10, 14], the results show that all the tested systems are vulnerable even against the crudest attacks such as high compressed images or images downloaded from social media websites. Thus, there is an urgent need to develop solutions to detect these spoofing attacks and protect face recognition systems.

Many face anti-spoofing methods have recently been proposed, a literature review of these methods can be found in [7]. Based on the printing defects and the artifacts caused by the lighting reflectance from the planar spoofing medium, the texture-based methods provide a sufficient mean to differentiate between real and fake face images. Most of these methods applied a single scale descriptor to encode the texture differences between the real and the fake faces. However, as can be seen in Figure 1, a face image can be of different image resolutions and qualities. As a consequence, a single scale descriptor cannot handle well all these different image resolutions and yields in unsatisfactory results. To deal with this problem, Määttä et al. [12] applied a multiscale descriptor (Multiscale Local Binary Pattern, MLBP) instead of single scale descriptor (LBP).

Figure 1: Face images from the CASIA Face Anti-Spoofing Database [27] illustrating the variation of the image resolution and quality (low, medium and high, respectively). Note that the original images have been rescaled to the same size.

In the present work, we propose to address the problem with image resolution and quality from a novel point of view. Instead of using multiscale descriptors, we propose to first represent the face images in different scales via multiscale filtering and then extract texture features from the scaled images. Besides encoding the face images in different scales, multiscale filtering can also act as a preprocessing to enhance the robustness of the face representation against factors such as noise and illumination. We

978-1-5090-1869-7/16/$31.00 ©2016 IEEE

particularly considered three type of multiscale filtering approaches namely Gaussian scale space, Difference of Gaussian scale space and Multiscale Retinex. To extract the texture features in the scaled images, we experimented with the popular Local Binary Patterns (LBP) texture operator [13]. Extensive experiments on three challenging databases, namely CASIA Face Anti-Spoofing [26], Replay-Attack[4] and MSU Mobile Face Spoof [21] databases clearly shows the effectiveness of our proposed approach compared to previous works.

2. Multiscale filtering techniques

Multiscale filtering analysis is motivated by the need to detect and characterize the edges of small and large structures alike. In a given image, different structures give rise to edges of varying scales: small scales correspond to fine details and large scales correspond to gross structures. In order to detect all image edges, image processing at different scales is thus needed.

Several multiscale filtering approaches have been proposed in the literature [19]. Among all the proposed techniques, those based on Gaussian kernels have very attractive properties partially inherited from the Gaussian function. These include linearity and spatial shift invariance. Moreover, structures at coarse scales can be easily related to those at finer scales and no new structures are created after smoothing.

In this work, we use three kinds of Gaussian based multiscale filtering techniques: Gaussian scale space (multiscale low-pass filtering), Difference of Gaussian (DoG) scale space (multiscale band-pass filtering) and Multiscale Retinex method (multiscale high-pass filtering). Descriptions of these three methods are given in the following sections.

2.1. Gaussian scale space

In the Gaussian scale space [23], a set of Gaussian kernels with different standard deviations are convolved with the original image to create images in different scales:

$$I(x, y, \sigma_i) = I(x, y) * G(x, y, \sigma_i), \tag{1}$$

where $*$ is the convolution operator and $G(x, y, \sigma_i)$ is the Gaussian kernel with the standard deviation σ_i defined by $G(x, y, \sigma_i) = \frac{1}{2\pi\sigma^2} \exp - \frac{x^2+y^2}{2\sigma_i^2}$.

The convolution with the Gaussian kernels can reduce the noise in the images and emphasize the coarser structures, which is desirable in texture analysis as textures are scale dependent.

2.2. Difference of Gaussian (DoG) scale space

The DoG filtering is an edge enhancement technique. It is obtained by subtracting a blurred image from another image. The blurred images are obtained by convolving the original image with Gaussian kernels with different standard deviations. Convolving an image with a Gaussian kernel suppresses the high-frequency spatial information while subtracting a blurred image from another preserves only the spatial information. Thus, the DoG filter is equivalent to a band-pass filter.

Given an image I, the DoG image at the scale s defined by the two parameters σ_i and σ_j is given by:

$$I(x, y, \sigma_i, \sigma_j) = (G(x, y, \sigma_i) - G(x, y, \sigma_j)) * I(x, y). \tag{2}$$

2.3. Multiscale Retinex

An image I can be seen as a multiplication between an illumination component L and a reflectance component R:

$$I(x, y) = L(x, y) \times R(x, y) \tag{3}$$

Since illumination can be considered as gradual changes, it corresponds to the lower frequencies in the spectrum. In contrast, the reflectance varies quite rapidly, so it can be considered as high frequencies in the spectrum. For an efficient image texture representation, the illumination component should preferably be reduced or removed. In the Multiscale Retinex method [9], the illumination component is estimated at each scale s by applying Gaussian low-pass filtering on the original image. Then, the reflectance component R at that scale is obtained by subtracting the log of the estimated illumination from the log of the original image I:

$$R(x, y, \sigma_i) = log(I(x, y)) - log[G(x, y, \sigma_i) * I(x, y)]. \tag{4}$$

3. Facial texture representation

The Local Binary Pattern (LBP) is used to extract the texture features from the different scale space images. A brief description of the LBP features and proposed multiscale facial texture representation is given in following.

3.1. Local Binary Patterns (LBP)

The LBP descriptor proposed by Ojala et al. [13] is a highly discriminative gray-scale texture descriptor. For each pixel in an image, a binary code (i.e. LBP pattern) is computed by thresholding a circularly symmetric neighborhood with the value of the central pixel. Finally, a histogram is created to collect the occurrences of different binary patterns. The LBP pattern of a pixel (x, y) extracted from the image (I) can be written as follows:

$$LBP_{P,R}(x, y) = \sum_{n=1}^{P} \delta(r_n - r_c) \times 2^{n-1}, \tag{5}$$

where $\delta(x) = 1$ if $x >= 0$, otherwise 0. r_c and $r_n (n = 1, ..., P)$ denote the intensity values of the central pixel

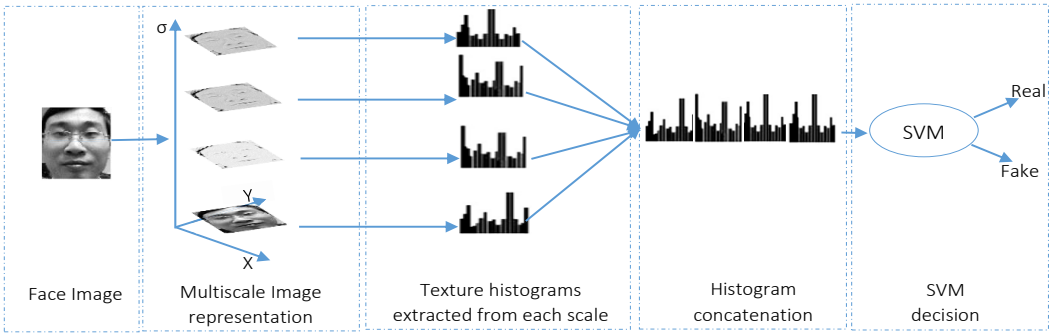

Figure 2: Illustration of the proposed face anti-spoofing approach.

(x, y) and its P neighborhood pixels located at the circle of radius R ($R > 0$), respectively. Another extension to the original LBP descriptor is the definition of the uniform patterns. A LBP pattern is defined as uniform if it's binary code contains at most two bitwise transition from 0 to 1 and vice versa [13].

3.2. Multiscale facial texture description

The pipeline of our proposed approach is illustrated in Figure 2. To extract the texture features from the different scale space images, the LBP descriptor is first applied on each scale image separately and then the resulting histograms are concatenated to form the final feature vector. Finally, the texture representation of a face image is fed into a binary classifier that determines whether the captured biometric train is originating from a genuine or a fake face.

4. Experiment data and setup

In this section, we first describe briefly the three benchmark datasets and the setup that was used in our experiments.

4.1. Experiment data

To assess the effectiveness of our proposed anti-spoofing technique, we considered three latest and most challenging publicly available face anti-spoofing databases: CASIA Face Anti-Spoofing Database (CASIA FASD), Replay-Attack Database and MSU Mobile Face Spoof Database (MSU MFSD).

4.1.1 CASIA Face Anti-Spoofing Database (CASIA FASD)

The CASIA Face Anti-Spoofing Database [27] contains video recordings of genuine and fake faces. The real faces were recorded from 50 genuine subjects, whereas the fake faces were made from the high quality recordings of the genuine faces. Three fake face attacks were designed: *warped photo attacks*, i.e. facial motion simulated by bending (warping) a photograph, *cut photo attacks* (photographic masks), i.e. the eye regions were cut off and the attacker hides behind the mask and exhibits eye-blinking through the holes, and *video attacks*. Both of real accesses and attacks attempts were recorded using three imaging qualities: *low*, *normal*, and *high*. The 50 subjects were divided into two subject-disjoint subsets for training and testing (20 and 30, respectively).

4.1.2 Replay-Attack Database

The Replay-Attack Database [4] consists of video recordings of real accesses and attack attempts to 50 clients. Using a built-in camera of a MacBook Air 13-inch laptop, a number of videos were recorded of each person in the database under two illumination conditions: *controlled*, i.e. uniform background and a fluorescent lamp was used to illuminate the scene, and *adverse*, i.e. non uniform background and the day-light was the only source of illumination. Under the same conditions a high resolution pictures and videos were taken for each person using a Canon PowerShot SX150 IS camera and an iPhone 3GS camera. These recordings were used to generate the fake face attacks.

Three types of attacks were designed: (1) *print attacks*, i.e. high resolution pictures were printed on A4 paper and displayed to the camera); (2) *mobile attacks*, i.e. high resolution pictures and videos were displayed on the iPhone 3GS screen; and (3) *high definition attacks*, i.e. the pictures and the videos were displayed on an iPad screen with resolution of 1024 by 768 pixels. The 50 subjects were divided on three subject-disjoint subsets for training, development and testing.

4.1.3 MSU Mobile Face Spoof Database (MSU MFSD)

The MSU Mobile Face Spoof Database [22] consists of 280 video recordings of real and fake faces. These recordings were taken from 35 subjects using two types of cameras: a

built-in camera of MacBook Air 13-inch laptop and a front-facing camera of a Google Nexus 5 Android phone. For the real accesses, each subject has two video recordings captured with the laptop and the Android cameras. For conducting the fake face video-replay attacks, first a high definition videos were taken for each subject using a Canon 550D single-lens reflex camera and an iPhone 5S back facing camera. The videos taken with the Canon camera were then replayed on iPad Air screen to generate the HD replay attacks while the videos recorded by the iPhone 5S mobile were replayed on the same device (the iPhone 5S mobile) to generate the mobile replay attacks. For the printed attacks, a HD pictures of the subject's faces were taken which then were printed on A3 paper using a HP colour laserjet CP6015xh printer. For the performance evaluation, the 35 subjects were divided on two subject-disjoint subsets for the training and testing (15 and 20 subjects, respectively).

4.2. Experiment setup

To ensure a fair comparison with the previous methods proposed in the literature, we followed the official overall test protocols associated with the three benchmark databases. For CASIA FASD and MSU MFSD, the model parameters are trained and tuned using a subject-disjoint cross-validation on the training set and the results are reported in terms of Equal Error Rate (EER) on the test set. The Replay-Attack Database provides also a separate development set for tuning the model parameters. Thus, the results are given in terms of EER on the development set and the Half Total Error Rate (HTER) on the test set following the official test protocol.

To be consistent with many previous works, all texture descriptions were extracted from face images of size 64×64 pixels. For the LBP descriptor, we used the $\text{LBP}^{u2}_{8,1}$ (i.e. $P = 8$ and $R = 1$) operator to extract the micro texture patterns. For the multiscale filtering methods, four scales (original image + three filtered images) were used to represent the face images. For the Gaussian scale space and the Multiscale Retinex methods, the filtered images are obtained using the standard deviation values: $\sigma_i = \{0.5, 1, 2\}$, while for the DoG scale space, the images are obtained using $(\sigma_1, \sigma_2) = \{(0.5, 1); (1, 2); (0.5, 2)\}$. In the following experiments, a linear Support Vector Machine (LIBLINEAR [5]) is used for classifying the feature vectors of each video frame separately.

The baseline method using multiscale LBP descriptor (MLBP) [12], was constructed by computing holistic description of the face with the $\text{LBP}_{8,1}$, $\text{LBP}_{8,2}$ and $\text{LBP}_{16,2}$ operators and concatenating the resulting histograms.

5. Results and discussion

In the first step of our experiments, we compute the performances of the LBP descriptor with and without the different multiscale filtering methods. The results are presented in Table 1. It is easy to see that the use of the multiscale filtering methods yield in significant performance improvement compared to the two baseline methods, i.e. based on LBP and MLBP descriptors. When comparing the performance of the three multiscale filtering techniques, we can observe that the DoG scale space gives the best results. The DoG scale space filtering improves the performance of the LBP descriptor on CASIA FASD, MSU MFSD and Replay-Attack Database with 34.8%, 25.8% and 32.3%, respectively.

The excellent performance of the DoG scale space can be explained by its ability to analyze the image properties at different frequency bands and minimizing the effect of noise and illumination. Gaussian scale space and Multiscale Retinex methods, on the other hand, are focusing only on the low frequencies or the high frequencies, respectively. In face anti-spoofing, it is important to analyze the face images at different frequency bands because we don't have prior knowledge about which frequencies contain the most discriminative inherent disparities between the real and the fake face images.

To show the improvement of the DoG filtering method on different operating points, we plotted the detection error trade-off (DET) curves of the LBP descriptor on the three databases in Figure 3. It can be seen that the DoG filtering based multiscale space representation perform consistently better than the other feature descriptions almost at all operating points on all three datasets.

Table 1: The performance in term of EER (%) and HTER (%) of the gray-scale LBP descriptor on CASIA FASD, MSU MFSD and Replay-Attack databases

Method	CASIA	MSU	Replay-Attack	
	EER	EER	EER	HTER
LBP	25.4	32.6	22.3	19.0
MLBP	17.9	29.6	23.2	22.2
Gausssian scale space + LBP	19.5	29.0	20.6	21.9
Multiscale Retinex + LBP	19.1	24.9	15.9	13.5
DoG scale space + LBP	**16.5**	**24.2**	**13.9**	**12.8**

In [3], the LBP features extracted from the color images show a very good performance improvement compared to those extracted from the gray-scale images. To improve further these performances we have applied the Difference of Gaussian scale space filtering on each channel of the RGB, HSV and YCbCr color spaces. The scale space LBP features extracted from each channel are then concatenated to from the final feature vector. A detailed description of these color spaces can be found e.g. in [17].

The results with and without the DoG scale space filtering are presented in the Table 2. We can clearly see that the

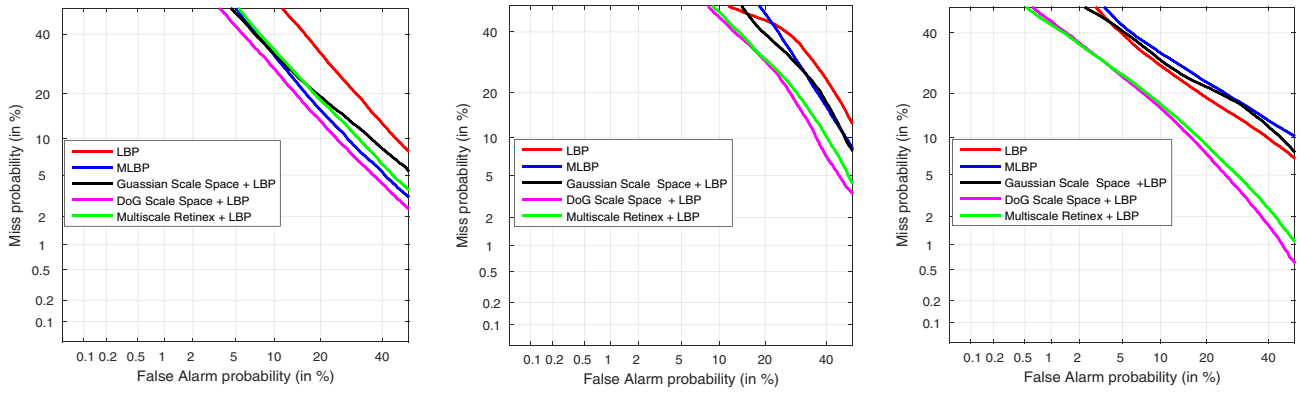

Figure 3: DET curves of the LBP descriptor on: a) CASIA FASD, b) MSU MFSD and c) Replay-Attack Database.

DoG scale space filtering improve the performance of the color LBP descriptor on the three color spaces and also in the combination of HSV and YCbCr color spaces (obtained by concatenating the texture features extracted from the two color spaces).

Table 2: The performance in term of EER (%) and HTER (%) of the color LBP descriptor on CASIA FASD, MSU MFSD and Replay-Attack databases with and without the DoG scale space filtering

Method	CASIA	MSU	Replay-Attack	
	EER	EER	EER	HTER
Without DoG scale space filtering				
RGB	19.1	15.6	5.9	9.4
HSV	12.6	17.2	4.3	7.7
YCbCr	13.2	11.1	4.3	7.4
HSV+YCbCr	7.1	10.6	0.9	4.9
With DoG scale space filtering				
RGB	10.7	11.7	5.1	5.9
HSV	8.9	11.5	3.2	8.5
YCbCr	8.8	9.4	3.6	9.6
HSV+YCbCr	**4.2**	**6.9**	**0.7**	**3.1**

Table 3 provides a comparison with the state-of-the-art face spoof detection techniques proposed in the literature. It can be seen that our proposed gives the state-of-the-art performance on the CASIA FASD and MSU MFSD, and very competitive results on the Replay-Attack Database. More importantly, unlike many of the methods proposed in the literature, our proposed approach is able to achieve stable performance across all the three benchmark datasets.

6. Conclusion

In face anti-spoofing, the face images can be captured at different image resolutions and qualities in different illumination conditions. This poses a serious problem to all texture-based face anti-spoofing methods. To deal with this problem, we presented a novel scale space texture analysis method. First, we applied multiscale filtering techniques to represent the original face images in a scale space representation. Then, using simple texture descriptor, we extracted texture features from each scale space image. The concatenated feature descriptions were used to differentiate fake face images from genuine ones. The extensive experiments on three most challenging face anti-spoofing benchmark databases pointed out the validity of our approach

Table 3: Comparison between the proposed countermeasure and state-of-the-art methods on the three benchmark datasets

Method	Replay-Attack		CASIA	MSU
	EER	HTER	EER	EER
DoG [27]	-	-	17.0	
Motion mag+LBP [2]	**0.2**	**0.0**	14.4	-
Noise signatures [16]	-	14.2	-	-
DMD [20]	5.3	3.7	21.7	-
Motion [1]	11.6	11.7	26.6	-
LBP [4]	13.9	13.8	18.2	
LBP-TOP [15]	7.9	7.6	10.0	-
CDD [25]	-	-	11.8	-
IQA[6]	-	15.2	32.4	-
CNN [24]	6.1	2.1	7.4	-
IDA [22]	-	7.4	-	8.5
Motion+LBP [8]	4.5	5.1	-	-
Colour LBP [3]*	0.9	4.9	7.1	10.6
Proposed method	0.7	3.1	**4.2**	**6.9**

* We reproduced the method proposed in [3] in order to get its frame based performance and results also on the new MSU dataset.

and particularity highlighted the high performance of the DoG scale space image representation. Strongly believing in the importance of reproducible research, we plan to make our implementation code publicly available to the research community upon the publication of this work.

Acknowledgments

The financial support of the Academy of Finland, Infotech Oulu, Nokia Foundation, the Northwestern Polytechnical University, and the Shaanxi Province is acknowledged.

References

[1] A. Anjos and S. Marcel. Counter-measures to photo attacks in face recognition: a public database and a baseline. In *Proceedings of IAPR IEEE International Joint Conference on Biometrics (IJCB)*, 2011. 5

[2] S. Bharadwaj, T. I. Dhamecha, M. Vatsa, and S. Richa. Computationally efficient face spoofing detection with motion magnification. In *Proceedings of IEEE Conference on Computer Vision and Pattern Recognition, Workshop on Biometrics*, 2013. 5

[3] Z. Boulkenafet, J. Komulainen, and A. Hadid. Face anti-spoofing based on color texture analysis. In *IEEE International Conference on Image Processing (ICIP)*, pages 2636 – 2640, 2015. 4, 5

[4] I. Chingovska, A. Anjos, and S. Marcel. On the effectiveness of local binary patterns in face anti-spoofing. In *International Conference of the Biometrics Special Interest Group (BIOSIG)*, pages 1–7, Sept 2012. 2, 3, 5

[5] R.-E. Fan, K.-W. Chang, C.-J. Hsieh, X.-R. Wang, and C.-J. Lin. Liblinear: A library for large linear classification. *J. Mach. Learn. Res.*, 9:1871–1874, June 2008. 4

[6] J. Galbally and S. Marcel. Face anti-spoofing based on general image quality assessment. In *Proc. IAPR/IEEE Int. Conf. on Pattern Recognition, ICPR*, pages 1173–1178, 2014. 5

[7] J. Galbally, S. Marcel, and J. Fiérrez. Biometric antispoofing methods: A survey in face recognition. *IEEE Access*, 2:1530–1552, 2014. 1

[8] J. Komulainen, A. Anjos, A. Hadid, S. Marcel, and M. Pietikäinen. Complementary countermeasures for detecting scenic face spoofing attacks. In *IAPR International Conference on Biometrics, ICB*, June 2013. 5

[9] E. H. Land. An alternative technique for the computation of the designator in the retinex theory of color vision. *Proceedings of the National Academy of Sciences*, 83(10):3078–3080, 1986. 2

[10] Y. Li, K. Xu, Q. Yan, Y. Li, and R. H. Deng. Understanding osn-based facial disclosure against face authentication systems. In *Proceedings of the 9th ACM Symposium on Information, Computer and Communications Security*, ASIA CCS '14, pages 413–424, New York, NY, USA, 2014. ACM. 1

[11] C. Lu and X. Tang. Surpassing human-level face verification performance on lfw with gaussianface. *arXiv preprint arXiv:1404.3840*, 2014. 1

[12] J. Määttä, A. Hadid, and M. Pietikäinen. Face spoofing detection from single images using micro-texture analysis. In *Proceedings of IAPR IEEE International Joint Conference on Biometrics (IJCB), Washington DC, USA*, 2011. 1, 4

[13] T. Ojala, M. Pietikäinen, and T. Mäenpaa. Multiresolution gray-scale and rotation invariant texture classification with local binary patterns. *IEEE Transactions on Pattern Analysis and Machine Intelligence (PAMI)*, 24(7):971–987, Jul 2002. 2, 3

[14] L. Omar and I. Ivrissimtzis. Evaluating the resilience of face recognition systems against malicious attacks. In G. K. L. Tam, editor, *Proceedings of the 7th UK Computer Vision Student Workshop (BMVW)*, pages 5.1–5.9. BMVA Press, September 2015. 1

[15] T. d. F. Pereira, J. Komulainen, A. Anjos, J. M. De Martino, A. Hadid, M. Pietikäinen, and S. Marcel. Face liveness detection using dynamic texture. *EURASIP Journal on Image and Video Processing*, 2013. 5

[16] A. Pinto, W. Robson Schwartz, H. Pedrini, and A. De Rezende Rocha. Using visual rhythms for detecting video-based facial spoof attacks. *Information Forensics and Security, IEEE Transactions on*, 10(5):1025–1038, May 2015. 5

[17] K. N. P. Rastislav Lukac. *Color Image Processing: Methods and Applications*, volume 8. New York CRC, 2007. 4

[18] Z. Stan and J. Anil. *Handbook of Face Recognition (second edition)*. Springer, 2011. 1

[19] J.-L. Starck, F. D. Murtagh, and A. Bijaoui. *Image processing and data analysis: the multiscale approach*. Cambridge University Press, 1998. 2

[20] S. Tirunagari, N. Poh, D. Windridge, A. Iorliam, N. Suki, and A. T. S. Ho. Detection of face spoofing using visual dynamics. *IEEE Transactions on Information Forensics and Security*, 10(4):762–777, 2015. 5

[21] D. Wen, H. Han, and A. Jain. Face Spoof Detection with Image Distortion Analysis. *IEEE Trans. Information Forensic and Security*, 10(4):746–761, April 2015. 2

[22] D. Wen, H. Han, and A. Jain. Face spoof detection with image distortion analysis. *Transactions on Information Forensics and Security*, 10(4):746–761, 2015. 3, 5

[23] A. P. Witkin. Scale-space filtering: A new approach to multi-scale description. In *Acoustics, Speech, and Signal Processing, IEEE International Conference on ICASSP '84.*, volume 9, pages 150–153, Mar 1984. 2

[24] J. Yang, Z. Lei, and S. Z. Li. Learn convolutional neural network for face anti-spoofing. *CoRR*, abs/1408.5601, 2014. 5

[25] J. Yang, Z. Lei, S. Liao, and S. Z. Li. Face liveness detection with component dependent descriptor. In *IAPR International Conference on Biometrics, ICB*, June 2013. 5

[26] Z. Zhang, J. Yan, S. Liu, Z. Lei, D. Yi, and S. Li. A face antispoofing database with diverse attacks. In *International Conference on Biometrics (ICB)*, pages 26–31, March 2012. 2

[27] Z. Zhang, J. Yan, S. Liu, Z. Lei, D. Yi, and S. Z. Li. A face antispoofing database with diverse attacks. In *5th IAPR International Conference on Biometrics (ICB)*, pages 26–31, 2012. 1, 3, 5