On the generalization of color texture-based face anti-spoofing

Zinelabidine Boulkenafet[a,*], Jukka Komulainen[a], Abdenour Hadid[a,b]

[a] *Center for Machine Vision and Signal Analysis, University of Oulu, Finland*
[b] *School of Electronics and Information at Northwestern Polytechnical University, Xian, China*

Abstract

Despite the significant attention given to the problem of face spoofing, we still lack generalized presentation attack detection (PAD) methods performing robustly in practical face recognition systems. The existing face anti-spoofing techniques have indeed achieved impressive results when trained and evaluated on the same database (i.e. intra-test protocols). Cross-database experiments have, however, revealed that the performance of the state-of-the-art methods drops drastically as they fail to cope with new attacks scenarios and other operating conditions that have not been seen during training and development phases. So far, even the popular convolutional neural networks (CNN) have failed to derive well-generalizing features for face anti-spoofing. In this work, we explore the effect of different factors, such as acquisition conditions and presentation attack instrument (PAI) variation, on the generalization of color texture-based face anti-spoofing. Our extensive cross-database evaluation of seven color texture-based methods demonstrate that most of the methods are unable to generalize to unseen spoofing attack scenarios. More importantly, the experiments show that some facial color texture representations are more robust to particular PAIs than others. From this observation, we propose a face PAD solution of attack-specific countermeasures based solely on color texture analysis and investigate how well it generalizes under display and print attacks in different conditions. The evaluation of the method combining attack-specific detectors on three benchmark face anti-spoofing databases showed remarkable generalization ability against display attacks while print attacks require still further attention.

Keywords: Face recognition, presentation attack detection, spoofing, color texture analysis, cross-database, generalization.

1. Introduction

The vulnerability of biometric systems against the learned or forged biometric traits has been the subject of many recent studies, including [1, 2, 3, 4, 5]. These works have concluded that most of the biometric systems, even those presenting a high recognition performance, are vulnerable to spoofing attacks (or a presentation attack as defined in the current ISO/IEC 30107-3 standard [6]). Face recognition systems are particularly easy to be deceived. With the increase of the social networks' popularity and the improvement of the camera resolution, it is easy to spoof the identity of a target person by using his/her images published in the web or captured from distance without permission. For instance, in a recent study [5], six commercial face recognition systems, namely Face Unlock, Facelock Pro, Visidon, Veriface, Luxand Blink and FastAccess, were easily fooled with crude photo attacks using images of the targeted person downloaded from social networks. Even worse, also their dedicated challenge-response based liveness detection mechanisms were circumvented using simple photo manipulation to imitate the requested facial motion (liveness cues), including eye blinking and head rotation.

To discriminate between real and fake face images, many face presentation attack detection (PAD) methods have been proposed in the literature (see [7, 8, 9] for extensive surveys). The existing face anti-spoofing techniques analyzing motion, facial texture content and image quality have already achieved impressive results particularly when trained and evaluated on the same database (i.e. intra-test protocols). As all the existing benchmark publicly available datasets lack variations in the collected data, e.g. user demographics, application scenarios, illumination conditions and input cameras, the reported anti-spoofing results may unfortunately not reflect the real uncontrolled operating conditions that will be definitely faced in real-world applications, such as mobile authentication. For instance, in the widely used Replay-Attack Database [10], the video samples of the training, development and test sets have been collected using a single camera.

To gain insight into the generalization performance of face anti-spoofing techniques, de Freitas Pereira *et al.* [11] suggested a cross-database evaluation in which the anti-spoofing models are trained and tuned on one database and then tested on other databases. The experiments have revealed that the performance of the state-of-the-art methods drastically drops as they failed to cope with new spoofing scenarios that have not been seen during training and development phases. So far, even the popular convolutional neural networks (CNN) have failed to derive well-generalizing features for face anti-spoofing [12, 13].

Cross-database testing has been increasingly applied in face PAD research [12, 13, 14, 15, 16, 17, 18, 19, 20] to overcome the shortcomings of the public datasets since the generalization

*Corresponding author
Email address: zinelabidine.boulkenafet@oulu.fi (Zinelabidine Boulkenafet)

issue was pointed out by de Freitas Pereira *et al.* [11]. This has been a nice trend but the main limitation with these preliminary studies has been that, in general, the generalization performance has been only broadly evaluated on the plain overall protocol (i.e. combining all types of spoofing scenarios) without any deep analysis on the effect of different factors such as input sensor or presentation attack instrument (PAI) variation on the generalization capability. Since the overall performance of the state of the art has been far from satisfying the strict security demands of biometric systems, one can even question the meaningfulness of this kind of benchmarking.

In this work, we show that the blind overall assessment might actually lead to overly pessimistic conclusions on the contrary as a method might be able to generalize under some conditions even if its plain overall performance is poor. We argue that careful breakdown analysis across different covariates, especially attack scenarios, is very crucial to gain better insights into the performance and importantly the generalization of different face anti-spoofing methods. The recently standardized ISO/IEC 30107-3 metrics [6] are an important step to the right direction because the attack potential is taken into account as the overall PAD performance corresponds to the most successful PAI. However, this indicates how easy a biometric system is to fool on average by exploiting its (possible) vulnerability, which suits well for evaluating the robustness of complete biometric solutions. Since it is reasonable to assume that no single superior technique is able to detect all known, let alone unseen, attacks types, it is also important to find out the operating conditions of different PAD methods and how complementary countermeasures could be combined to achieve more robust overall performance [21].

Based on the above observations, we present in this work an in-depth analysis on the generalization of color texture-based face anti-spoofing. This is motivated by our recent works [16, 17, 18] showing that color texture features extracted from both luminance and chrominance color channels provide the state-of-the-art performance and very promising generalization abilities in face PAD. We perform extensive cross-database tests which measure the robustness of seven different facial color texture descriptions across different covariates, like acquisition conditions and attack scenarios. Our experiments depict that most of methods are unable to generalize to unseen spoofing attack scenarios but some of the methods are more robust to particular PAIs. Inspired by this, we propose an attack-specific approach to cope with the problem of generalized face PAD. Compared to the state of the art, we obtained very competitive intra-database and inter-database results on three benchmark face spoofing databases. More importantly, the color texture based method can generalize extremely well against display attacks (digital photo and video-replay attacks) launched at short distance, while further work or other complementary countermeasures is needed for tackling print attacks.

The rest of the article is organized as follows. First, in Section 2, we give a brief overview on the different approaches for face PAD proposed in the literature. Section 3 presents the different color texture descriptors studied in this work. The experimental setup is described in Section 4. Section 5 is devoted to the in-depth analysis, exploring the generalization problem across different conditions, and describing the newly proposed scheme along with a fair comparison against state of the art. Concluding remarks are drawn in Section 6.

2. Related work

There exists no unified taxonomy for the different face PAD approaches. In this article, we categorized the methods into two groups: hardware-based and software-based methods.

Hardware-based methods are probably the most robust ones for anti-spoofing because the dedicated sensors are able to directly capture or emphasize specific intrinsic differences between genuine and artificial faces in 3D structure [22, 23] and (multi-spectral) reflectance [23, 24, 25, 26] properties. For instance, planar PAI detection becomes rather trivial if depth information is available [22], whereas near-infrared or thermal cameras are efficient in display attack detection as most of the displays in consumer electronics emit only visible light. On the other hand, these kinds of unconventional sensors are usually expensive and not compact, thus not (yet) available in mobile devices, which prevents their wide deployment.

It would be rather appealing to perform face PAD by further analyzing only the same data that is used for the actual biometric purposes or additional data captured with the standard acquisition device. These kinds of software-based methods can be broadly divided into active (requiring user collaboration) and passive approaches. Additional user interaction can be very effectively used for face anti-spoofing because we humans tend to be interactive, whereas a photo or video-replay attack cannot respond to randomly specified action requirements. Furthermore, it is almost impossible to perform liveness detection or facial 3D structure estimation by relying only on spontaneous facial motion. Challenge-response based methods aim at performing face PAD detection based on whether the required action (challenge), e.g. facial expression [27, 28], mouth movement [27, 29] or head rotation (3D structure) [30, 31, 32], was observed within a predefined time window (response). Also, active software-based methods are able to generalize well across different acquisition conditions and attack scenarios but at the cost of usability due to increased authentication time and system complexity.

Ideally, passive software-based methods would be preferable for face PAD because they are faster and less intrusive than their active counterparts. Due to the increasing number of public benchmark databases, numerous passive software-based approaches have been proposed for face anti-spoofing. In general, passive methods based on analyzing different facial properties, like frequency content [33, 34], texture [10, 35, 36, 37, 38, 39] and quality [40, 41, 42], or motion cues, like eye blinking [43, 44, 45, 46], facial expression changes [27, 44, 45, 46], mouth movements [27, 44, 45, 46], or even color variation due to blood circulation (pulse) [47], to discriminate face artifacts from genuine ones. Passive software-based methods have shown impressive results on the publicly available datasets but the preliminary cross-database tests, like [11, 32], revealed that

the performance is likely to degrade drastically when operating in unknown conditions.

Recently, the research focus on software-based face PAD has been gradually moving into assessing and improving the generalization capabilities of the proposed and existing methods in a cross-database setup instead of operating solely on single databases. Among hand-crafted feature based approaches, image distortion analysis [14], combination of texture and image quality analysis with interpupillary distance (IPD) based reject option [19], dynamic spectral domain analysis [15] and pulse detection [48] have been applied in the context of generalized face anti-spoofing but with only moderate results.

The initial studies using deep CNNs have resulted in excellent intra-test performance but the cross-database results have still been unsatisfactory [12, 13]. This is mainly due to the fact that the current publicly available dataset may not probably provide enough data for training well-known deep neural network architectures from scratch or even for fine-tuning pre-trained networks, thus the CNN models have been suffering from overfitting. In [20], deep dictionary learning based formulation was proposed to mitigate the requirement of large amounts of training data with very promising intra-test results but the generalization capability was again unsatisfying. In order to exploit CNNs to their full potential, novel techniques for cross-domain adaptation are needed, or application-specific learning needs to be further explored when more comprehensive databases are available.

Cross-database testing has been indeed an important paradigm shift in the research community towards generalized face PAD. Unfortunately, so far the average overall performance of the best-performing method [13] across different databases has been 22% in terms of Half Total Error Rate (HTER). This demonstrates clearly that the generalization capabilities of the existing methods are not at acceptable level, thus they cannot be directly utilized in real-world applications. However, in this work, we show that some of the methods might be actually able to generalize reasonably well under some conditions, like particular attack types and scenarios. This can be easily overlooked as the poor overall cross-database performance does not really encourage to conduct a closer examination of the results.

3. Color texture descriptors for face PAD

For our generalization study, we considered color texture analysis based face PAD because it has shown promising generalization capabilities in our recent works [16, 17, 18]. The key idea behind color texture based face anti-spoofing is that an image of an artificial face is actually an image of a face which passes through two different camera systems and a printing system or a display device, thus it can be referred to in fact as a recaptured image. As a consequence, the observed artificial face image is likely to suffer from different kinds of quality issues, such as printing defects, video artifacts, PAI dependent (local) color variations and limited color reproduction (gamut), that can be captured by analyzing the texture content of both luminance and chrominance channels.

We have shown in our previous works [16, 17, 18] that extracting texture features separately from luminance and chrominance channels of the HSV and YCbCr color spaces provide efficient and complementary facial color texture descriptions for anti-spoofing because the representation of chroma components in the two color spaces is different. The three color components of RGB color space (red, green and blue) are highly correlated, while both HSV and YCbCr color spaces are based on the separation of the luminance and the chrominance components. In the HSV colour space, hue and saturation define the chrominance of the image while value corresponds to the luminance. The YCbCr space separates the RGB components into luminance (Y), chrominance blue (Cb) and chrominance red (Cr). For further details on the color texture based face PAD, interested readers are referred especially to [17], while more information about different color spaces can be found e.g. in [49].

In this present work, we selected seven feature descriptors for extracting the facial color texture representation from the HSV and YCbCr color spaces. These descriptors include Uniform Local Binary Patterns (LBP), Rotation Invariant Uniform Local Binary Patterns (RI-LBP), Binarized Statistical Image Features (BSIF), Co-occurrence of Adjacent Local Binary Patterns (CoALBP), Rotation Invariant Co-occurrence among Adjacent Local Binary Patterns (RIC-LBP), Local Phase Quantization (LPQ) and Speed-Up Robust Features (SURF). As some of these texture descriptors were originally designed to operate on gray-scale images, they are adapted to analyze color images by combining the texture features extracted from different color channels. A short description of these descriptors is given in the following.

3.1. Local Binary Patterns (LBP)

The Local Binary Patterns descriptor is a discriminative gray-scale texture descriptor proposed by Ojala *et al.* [50]. For each pixel in an image, a binary code is computed by thresholding a circularly symmetric neighborhood with the value of the central pixel.

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

where $\delta(x) = 1$ if $x >= 0$, otherwise $\delta(x) = 0$. r_c and $r_n (n = 1, ..., P)$ denote the intensity values of the central pixel (x, y) and its P neighborhood pixels located at the circle of radius R ($R > 0$), respectively. To represent the image texture information, the occurrences of the different binary codes are collected into a histogram.

The uniform LBP (LBP^{u2}) and the rotation invariant uniform LBP (LBP^{riu2}) are two extensions of the LBP operator. An LBP pattern is considered as uniform if its binary code contains at most two transitions from 0 to 1 or from 1 to 0. In the LBP^{riu2} descriptor, the LBP^{u2} binary code is shifted until it corresponds to one of the pre-selected rotation invariant patterns.

3.2. Co-occurrence of Adjacent Local Binary Patterns (CoA-LBP)

In the original LBP method, the collection of the LBP patterns into one histogram discard the spatial relation between the

patterns. To exploit this spacial information, the Co-occurrence of Adjacent LBP (CoA-LBP) method [51] was proposed. In this method, first, the LBP patterns are exacted from the images using the simplified LBP descriptors (LBP_+ or $LBP_\times$). Then, to exploit the correlation between the adjacent patterns, four directions were defined: $D = \{(0, \Delta d), (\Delta d, 0), (\Delta d, \Delta d)$ and $(-\Delta d, \Delta d)\}$ where Δd is the distance between two adjacent LBP patterns. For each direction, a 16×16 2-D histogram is created then the resulting histograms are reshaped and concatenated to form the final feature descriptor. In the Rotation Invariant Co-occurrence LBP (RIC-LBP) [52], the CoA-LBP patterns corresponding to a pre-selected rotation invariant code are pooled together giving a reduced feature vector.

3.3. Local Phase Quantization (LPQ)

The Local Phase Quantization (LPQ) descriptor [53] was proposed to extract the texture information from the blurred images. It uses the Short Term Fourier Transform (STFT) to analyze the $M \times M$ neighborhoods surrounding a target pixel x. Let $F_u(x)$ be the output of the STFT at the pixel x using the bi-dimensional spatial frequency u. In the LPQ descriptor, only four complex frequencies are used: $u_0 = (\alpha, 0)$, $u_1 = (\alpha, \alpha)$, $u_2 = (0, \alpha)$, $u_3 = (-\alpha, -\alpha)$ where α is a small scalar ($\alpha << 1$). These frequencies correspond to the directions 0, 45, 90 and 135. The LPQ features at a pixel x are given by the vector $F_x = [Re\{F_{u_0}(x), F_{u_1}(x), F_{u_2}(x), F_{u_3}(x)\}, Im\{F_{u_0}(x), F_{u_1}(x), F_{u_2}(x), F_{u_3}(x)\}]$ where $Re\{.\}$ and $Im\{.\}$ are the real and the imaginary parts of a complex number, respectively. The elements of the vector F_x are binarized using the δ function defined previously then the resulting binary coefficients are represented as integer values in [0-255] and collected into a histogram. To make the LPQ coefficients statistically independents, a de-correlation step based on the whitening transform is suggested and applied before the quantization process.

3.4. Binarized Statistical Image Features (BSIF)

For each pixel in an image, the Binarized Statistical Image (BSIF) descriptor [54] computes a binary code string. Each bit, in this code, is computed by binarizing the response of a linear filter with a threshold at zero. Let X be an image patch of size $l \times l$ and let W_i be a filter of the same size. The response of the filter W_i is obtained by:

$$S_i = \delta(\sum_{u,v} W_i(u, v)X(u, v)) = \delta(w_i x), \tag{2}$$

where w_i and x are the vectors which contain the pixel values of W_i and X, respectively. The length of the binary code is determined by number of filters used. The filters' coefficients are learnt by maximizing the statistical independence of the filter responses using natural image patches.

3.5. The Speed Up Robust Features (SURF)

The Speed Up Robust Features (SURF) [55] is an interest point detector proposed to improve the speed of the Scale Independent Feature Transform (SIFT) descriptor. The SURF descriptor uses the Harr box filters to approximate the Laplacian of Gausssian instead of using the Difference of Gaussian (DoG) filters. The convolution with these box filters can be easily computed using the integral images and it can be paralyzed at different scales.

The region around each interest point is first divided into 4×4 sub-regions. Then, for each sub-region j, the horizontal and vertical Wavelet responses are used to form a feature vector V_j as follows:

$$V_j = [\sum d_x, \sum d_y, \sum |d_x|, \sum |d_y|] \tag{3}$$

The feature vectors extracted form each sub-region are then concatenated to from a $SURF$ descriptor with 64 dimensions.

$$SURF = [V_1, ..., V_{16}] \tag{4}$$

Before classification, we use the Fisher Vector (FV) method [56] to encode the SURF features. FV encoding embeds the SURF features in a high-dimensional space by fitting a generative parametric model (Gaussian Mixture Model GMM) to the features to be encoded. The encoded features represents how the distribution of the local descriptors differ from the distribution of the GMM model learnt with all the training images. The FV representation is normalized using a square rooting followed by L_2 normalization, which leads to excellent results even with efficient linear classifiers [56, 57]. To de-correlate the SURF features and reduce their dimensionality, Principal Component Analysis (PCA) is applied before the FV encoding.

Table 1: The parameters for the different descriptors and the dimensions of the concatenated HSV and YCbCr texture representations used in our experiments

Method	Parameters	Dimension
LBP	Radius R=1, Neighbors P=8	354
RI-LBP	Radius R=1, Neighbors P=8	60
LPQ	Widows size M=7 and α=1/7	1536
BSIF	Filter size l=7×7, Number of filters=8	1536
CoALBP	R=1, LBP descriptor= LBP_+, B=2	6144
RIC-LBP	R=1, LBP descriptor= LBP_+, B=2	408
SURF+FV	Step= 2 pixels, PCA=300, GMM=256	153600[1]

4. Experimental setup

In this study, the generic pipeline for performing color texture based face spoofing detection is as follows. First, the face is detected, cropped and geometrically normalized into a 64×64 pixel image based on the eye locations. The normalized RGB face image is converted into HSV and YCbCr color spaces and holistic texture features are extracted separately from each channel. The resulting feature vectors are concatenated into an enhanced feature vector in order to get an overall representation of the facial color texture. Finally, the obtained feature vectors are fed into a Softmax classifier. The score value describes

[1]The dimension of the SURF features before and after PCA is 384 and 300, respectively, while FV encoding embeds the de-correlated SURF features in a high-dimensional space more amenable to linear classification [18, 56, 57].

Table 2: Summary of three benchmark face presentation attack databases: Replay-Attack Database, CASIA FASD and MSU-MFSD .

Database	# subjects	Acquisition devices	# lighting scenarios	Attacks	# real/attack videos
Replay-Attack [10]	50	1 laptop	2	1 printer & 2 displays	200/1000
CASIA-FASD[14]	50	2 webcams & 1 compact system camera	1	1 printer & 1 display	150/450
MSU-MFSD [58]	35	1 laptop & 1 smartphone	1	1 printer & 2 displays	110/330

Table 3: The performance (HTER) of the different descriptors on the display vs display scenario

Train on:	CASIA		Replay		MSU		Average
Test on:	Replay	MSU	CASIA	MSU	CASIA	Replay	
LBP	29.1	33.7	26.4	23.9	32.3	30.2	29.3
RI-LBP	8.7	9.9	25.2	21.8	36.3	14.4	**19.4**
BSIF	35.4	17.1	37.7	34.2	42.6	47.6	35.8
LPQ	33.9	20.1	38.5	34.0	44.8	46.4	36.3
CoLBP	17.2	21.0	20.8	23.7	32.7	27.6	23.8
RIC-LBP	25.4	16.1	26.1	24.5	36.2	16.5	24.1
SURF	26.9	14.8	17.2	26.7	43.1	40.6	28.2

Table 4: The performance (HTER) of the different descriptors on the print vs print scenario

Train on:	CASIA		Replay		MSU		Average
Test on:	Replay	MSU	CASIA	MSU	CASIA	Replay	
LBP	30.9	51.5	41.5	43.0	56.7	48.3	45.3
RI-LBP	23.1	53.7	38.9	53.3	56.7	48.1	45.6
BSIF	29.6	48.4	32.5	49.1	50.0	49.4	43.2
LPQ	31.5	40.2	37.7	42.4	49.0	49.6	41.7
CoLBP	33.7	50.2	43.3	46.0	55.3	52.3	46.8
RIC-LBP	21.5	49.4	41.9	53.5	61.2	64.9	48.7
SURF	27.2	34.7	32.4	27.2	36.1	46.0	**34.0**

whether there is a live person or a fake one in front of the camera. The parameters and the feature dimensions of the different descriptors used in our experiments are provided in Table 1.

For our extensive experimental analysis, we considered three publicly available face anti-spoofing databases, namely CASIA Face Anti-Spoofing Database (CASIA FASD) [58], Replay-Attack Database [10] and MSU Mobile Face Spoof Database (MSU MFSD) [14]. These are the most challenging face anti-spoofing benchmark databases consisting of video recordings of real client accesses and various presentation attacks. These videos are captured with different imaging qualities, including mobile phones, webcams and (compact) system cameras. Table 2 provides a summary of the three databases in terms of number of subjects, lighting scenarios, printers and display devices used to create the attacks, and real and attack videos.

We followed the official test protocols of the three databases throughout our intra-database experiments, which allows a fair comparison with other methods proposed in the literature. Since the CASIA FASD and MSU MFSD lack a pre-defined development set, the model parameters are trained and tuned using 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 a separate development set for tuning the model parameters. Thus, the results on the test set are given in terms of HTER which is the average of False Acceptance Rate (FAR) and False Rejection Rate (FRR) at the decision threshold defined by the EER on the development set. Similarly, in our cross-database tests, we used the training set to build the countermeasure models and the test set to estimate the EER threshold which is applied on the other databases to compute the generalization performance in terms of HTER.

5. Experimental analysis

In this section, we present our in-depth analysis on the generalization of color texture based face PAD. We begin our experiments by comparing the robustness of the different color texture features to different PAIs. Then, we combine complementary facial color texture representations to form the final face description used in our anti-spoofing method and provide extensive experimental analysis and discussion. Finally, we compare our intra-database and cross-database performance against that of state-of-the-art algorithms.

5.1. Generalization across different PAIs

In the following experiments, we analyze the attack-specific generalization capabilities of the different color texture descriptors in cross-database setup. We considered two very broad classes of attacks based on the used PAI: print and display attacks (consisting of both digital photos and video-replays) because they are present in almost all existing face anti-spoofing databases, including the three databases used in our study. In each scenario, only one attack type (display or print) is present in training, development and test sets.

The results on the two scenarios: *display vs display* and *print vs print* are presented in Table 3 and Table 4, respectively. In general, the color texture based methods can generalize significantly better in detecting the display attacks compared to the print attacks. The simple and compact texture operator RI-LBP outperforms the other feature descriptors in the *display vs display* scenario with an overall HTER of 19.4% over all the databases. In the *print vs print* scenario, by far, the best results are obtained using the SURF features with an average HTER of 34.0%.

5.2. Fusion of attack-specific methods

The findings of the previous experiment confirmed our hypothesis that complementary attack-specific algorithms are probably needed for detecting different attack types. Even though cross-database results in the case of print attacks were not particularly good, we investigate next how well a face PAD

Table 5: The performance of our proposed method on the cross-database scenario

Train on:	CASIA		Replay		MSU		Average
Test on:	Replay	MSU	CASIA	MSU	CASIA	Replay	
RI-LBP	17.9	18.7	34.9	30.0	47.5	46.3	32.5
SURF	28.0	20.1	22.5	31.2	25.2	29.5	26.1
RI-LBP+SURF *	14.0	20.7	32.7	31.3	45.0	45.9	31.6
Proposed	9.6	19.8	39.2	33.3	29.7	21.4	**25.5**

* RI-LBP+SURF is trained with both display and print attacks then the resulting scores are fused using the simple sum method.

Table 6: The performance of the proposed method on the different type of attacks.

Train on:		CASIA		Replay		MSU		Average
Test on:		Replay	MSU	CASIA	MSU	CASIA	Replay	
All attacks	EER	9.9	19.9	31.2	21.4	29.2	16.2	21.3
	HTER	9.6	19.8	39.2	33.3	29.7	21.4	25.5
Display attacks	EER	6.5	7.6	29.7	13.6	27.4	10.1	15.8
	HTER	6.4	9.8	34.9	25.3	27.4	18.4	20.4
Print attacks	EER	22.8	33.0	33.2	50.3	30.0	33.7	33.9
	HTER	22.3	43.9	41.5	52.6	30.9	33.4	37.4

Table 7: The performance of the proposed method on the different types of attacks using combined training sets

Train on:		CASIA+MSU	CASIA+Replay	MSU+Replay	Average
Test on:		Replay	MSU	CASIA	
All attacks	EER	10.2	18.4	22.6	17.0
	HTER	9.6	19.0	22.8	17.2
Display	EER	7.9	6.0	27.5	13.8
	HTER	7.8	9.2	25.7	14.2
Print	EER	19.0	26.6	20.3	21.9
	HTER	18.7	42.6	21.5	27.6

solution based solely on color texture analysis generalizes under all sorts of attacks. Among the studied facial color texture descriptions, those based on RI-LBP and SURF features were clearly the most robust in the *display vs display* and *print vs print* scenarios, respectively. Therefore, we propose a face PAD scheme based on the fusion of the two complementary attack-specific facial color texture representations: one based on the RI-LBP descriptor for detecting the display attacks and the other based on the SURF descriptor for detecting the print attacks. The model based on the RI-LBP descriptor is trained using only real and display attack samples while the model based on the SURF descriptor is trained using only the real and the print attack samples. The countermeasure models based on RI-LBP and SURF features are combined at the score level using simple sum fusion rule.

The obtained cross-database results are summarized in Table 5. To gain insight into the significance of our proposed attack-specific fusion and training scheme, we have also reported the performance when both models were trained using both display and print attacks (RI-LBP+SURF). From these results, we can observe that the proposed method with specialized detectors yields in more robust performance on average than with the models trained using both PAIs. The breakdown analysis reported in Table 6 shows that the generalization performance against display attacks is much better compared to print attacks. These results confirm the benefits of attack-specific training as the proposed fusion scheme is able to maintain the performance of the two individual attack-specific detectors even when both PAIs are present in the test data.

The performance of the proposed method varies a lot between the different test scenarios. The main reason behind this is the amount of representative training data as the CASIA FASD contains more variations in the collected data, e.g. imaging quality, attack presentations and proximity between the camera and the face, compared to the Replay-Attack Database and MSU MFSD. Therefore, the model optimized for these databases has difficulties to perform well in the authentication scenarios. In order to study how the amount of training data affects the generalization performance, we combined two of the databases for training and used the remaining one for testing. As expected, the results presented in Table 7 demonstrate that the use of more comprehensive training set yields in more stable performance on both the Replay-Attack Database and the MSU MFSD and improves the generalization on the CASIA FASD.

5.3. Further analysis

In the following, we provide further discussion on: 1) the attack standoff distance, 2) differences between display and print attack detection, and 3) "standard" performance metrics used in face anti-spoofing literature and system tuning.

Table 8: The performance of the proposed method on the different imaging qualities of the CASIA FASD database: low, normal and high

Train on:		Replay			MSU			Replay+MSU		
Test on:		Low	Normal	High	Low	Normal	High	Low	Normal	High
All attacks	EER	31.2	36.6	25.3	31.8	30.4	18.7	21.3	22.2	11.1
	HTER	42.0	49.0	26.2	39.0	29.9	24.0	21.6	36.0	11.8
Display attacks	EER	29.2	36.5	19.3	31.8	31.0	16.2	26.1	32.1	3.8
	HTER	38.1	47.6	17.9	37.6	32.1	12.7	25.8	45.6	3.7
Print attacks	EER	32.4	35.8	31.8	31.8	29.7	19.3	19.4	16.1	12.0
	HTER	43.6	49.9	30.7	39.5	28.7	30.2	20.0	30.8	16.1

Figure 1: Sample images highlighting the variation of the standoff distance between the low, normal and high imaging qualities of the CASIA FASD. Please note that the original 1920×1080 resolution videos have been cropped into patches of 1280×720 pixels containing mainly the face region [59].

5.3.1. Attack standoff distance

Both of real and attack videos in the CASIA FASD were recorded using three imaging qualities: low (L), normal (N), and high (H). At first glance, it may seem that the results on the CASIA FASD database are still unsatisfying even under display attacks. However, a closer look at the results at the three different imaging qualities reported in Table 8 reveals that actually the proposed method trained on both Replay-Attack Database and MSU MFSD generalizes extremely well in the high imaging quality scenario, especially against display attacks. Similar observations can be made when the models are trained only on a single dataset.

The low and normal quality scenarios in the CASIA FASD are somewhat similar in the end. In principle, they both share the same video resolution (480×640 and 640×480, respectively) and the quality difference is mainly due to the age of the used webcam as aging degrades the imaging quality [59]. The high quality samples, however, have been collected using a Sony NEX-5 (compact) system camera with a resolution of 1920×1080 (cropped into patches of 1280×720 in order to save memory and computational burden [59]). As seen in Figure 1, the standoff distance between the attack presentations and the input sensor is much larger in both the low and normal imaging quality scenarios compared to the videos recorded with the high quality camera and, more importantly, also to the videos in the Replay-Attack Database and MSU MFSD. Since the cross-database results on the low and normal imaging qualities are similar and the video resolution is comparable to the ones used in the Replay-Attack Database and MSU MFSD, we believe that the lack of generalization on the CASIA FASD is mainly due to the difference in the standoff distance rather than sensor interoperability issues in this case.

Attack presentation can be performed with large or small standoff distance. Presentation attacks with large standoff are difficult to detect based on the facial texture information because the resulting face size (image resolution) might be too small to distiguish the PAI dependent artifacts. Fortunately, larger standoff means usually that contextual cues, like the a bezel (frame) of a display device or photograph edges, or the attackers' hands, might be visible in the provided view and easily detected with algorithms utilising the whole video frame for PAD, like [13, 19, 60]. The contextual visual cues can be concealed by placing the PAI very close to the input sensor but this may cause in defocus and results in larger resolution facial images that reveal PAI related characteristic quality degradations better. Our cross-database experiments using the combined training sets support this intuition as the proposed color texture based method is able to achieve promising overall generalization performance against attacks launched at short distance (9.6%, 19.0% and 11.8% in terms of HTER on the Replay-Attack Database, MSU MFSD and the H protocol of the CASIA FASD, respectively). Since the users aiming at positive verification can be expected to be cooperative, most of the commercial mobile face authentication systems are using assisted data capture mode for roughly fixing acquisition setup. Consequently, the same approach could be used for restricting the standoff distance, and potentially revealing also new visual cues, for PAD purposes.

5.3.2. Display vs print attack detection

In general, the main finding of our experiments is that the very promising generalization performances are obtained in dis-

play attack detection while only moderate detection results are achieved in the case of print attacks. This can be explained by taking a close look at: 1) the attack variation between the benchmark databases, 2) the inherent differences between display and print attacks in general.

The high quality display attacks in all three databases are performed using an iPad (presumably a 1st generation iPad in the Replay-Attack Database and the CASIA FASD and a 1st generation iPad Air in the MSU MFSD). The Replay-Attack Database and the MSU MFSD contain also mobile phone attacks launched using iPhone 3GS and iPhone 5S, respectively. Thus, it is likely that the display attacks are similar between the three databases. The print attacks, on the other hand, vary significantly between the three datasets because different printers and paper qualities are used for performing the attacks. In addition, the used paper size on the CASIA FASD and the Replay-Attack Database is A4, while A3 is used on the MSU MFSD. Another issue is that, in principle, print attacks have higher resolution (and quality) than display attacks due to the size of PAI. Therefore, print attacks can be presented from longer distance compared to display attacks. For instance, in MSU MFSD, the average standoff for iPhone 5S, iPad Air, A3 paper print attacks and genuine face are 10cm, 20cm and 40 cm and 50cm, respectively. Thus, the quality of the display attacks is lower compared to the print attacks whose standoff is almost as long as for real subjects.

Despite the fact that the display attacks might be similar between the different datasets, it is worth highlighting that the display attacks suffer from device-independent artifacts in general, thus have less textural variations in the recorded images. For instance, when the proposed method is trained only on CASIA FASD that does not contain mobile phone attacks, remarkable generalization is obtained on the display attacks of the Replay-Attack Database and MSU MFSD (6.4% and 9.8% in terms of HTER, respectively). Due to the very short standoff, the display attacks are likely to be defocused and low contrast, and moiré effects and other noise signatures are much more evident the closer a (high-quality) camera is to a display PAI. The proposed color texture based method is able to capture well these kinds of degradations if the attacks are launched at close distance. This can be attested by the astonishing generalization performance achieved in the cross-database test using combined training sets as the HTERs in display attack detection are 7.8%, 9.2% and 3.7% on the Replay-Attack Database, MSU MFSD and the H procotol of the CASIA FASD, respectively. The generalization against display attacks is indeed very promising but should be still treated with caution as the variation of display types between the existing bechmark datasets is limited.

Print attacks, on the other hand, have larger variations in the texture information as the use of different printers and paper qualities (and sizes) alter the inherent high-frequency artifacts, e.g. printing signatures. While the proposed approach might be suitable for capturing the display related artifacts, the studied color texture representations of the downsampled 64×64 face images are not optimal for describing the higher frequency type noise signatures of print attacks. Thus, color texture based print attack detection requires still further attention, or, alternatively, approaches utilizing e.g. motion, eyeblink or pulse detection could be coupled with color texture based methods in order to increase the robustness to print attacks.

5.3.3. Performance metrics and system tuning

When investigating more closely the scenarios in which the proposed color texture based approach failed to generalize, we computed the corresponding EER on the test set in addition to HTER. We noticed that in some cases there is a significant inconsistency between the EER and HTER. For instance from Table 6, in the case of display attacks, when the model trained and tuned on the Replay-Attack Database is tested on the MSU MFSD, the HTER is far from satisfactory, whereas the EER is actually quite good (25.3% and 13.6%, respectively). Also, when the model trained and tuned on the MSU MFSD is evaluated on the Replay-Attack Database, the HTER is again significantly higher than the EER (18.4% and 10.1%, respectively). Again, by looking only at the plain HTER values, one might overlook the fact that a method is actually able to generalize under some conditions. To be more specific, the color texture based method is actually able to separate the two classes (real vs display) well but the operating point, i.e. HTER threshold, is severely biased between the development and test sets. For instance, metric learning [61] might be one possible approach to consider for future work to mitigate this kind of tuning issues in transferring face PAD methods into practice.

5.4. Comparison against the state of the art

For the sake of completeness, Tables 9 and 10 provide a thorough comparison between our proposed face PAD scheme and the state-of-the-art methods. From the intra-test comparison presented in Table 9, it can be noticed that the proposed approach gives comparable performance on all three benchmark datasets. The results of the cross-database setup shown in Table 10 demonstrate the good overall generalization ability of our method compared to the state of the art. When the model is trained on the CASIA FASD and evaluated on the Replay-Attack Database, a significant improvement can be observed. Even the recent method combining eyeblink detection and a deep CNN framework analyzing both facial texture and contextual cues is left behind. In general, the color texture analysis seems to be promising direction for future face PAD studies as the state-of-the-art results in all scenarios are obtained using a color texture based method. It is also worth mentioning that our cross-database results obtained by training the model with combined databases are comparable to the intra-database performance of some existing methods, like gray-scale LBP.

6. Conclusion

The face anti-spoofing methods have shown astonishing results on the individual benchmark databases but have failed to generalize in more realistic setup introducing previously unseen acquisition conditions and attack types, for instance. In the preliminary cross-database studies, the generalization performance has been only broadly evaluated on the plain overall

Table 9: Comparison against state of the art (intra-database)

Method	Replay-Attack		CASIA	MSU
	EER	HTER	EER	EER
Motion [62]	11.6	11.7	26.6	-
LBP [10]	13.9	13.8	18.2	
DoG [58]	-	-	17.0	
LBP-TOP [63]	7.9	7.6	10.0	-
magLBP+magHOOF [45]*	0.0	0.2	14.4	-
LBP+HOOF [46]	-	-	3.1	**0.0**
CDD [37]	-	-	11.8	-
IQA[40]	-	15.2	32.4	-
Spectral cubes [15]	-	2.8	14.0	-
CNN [12]	6.1	2.1	7.4	-
DMD [44]	5.3	3.8	21.8	-
IDA [14]	-	7.4	-	8.5
Deep dictionary [20]	-	**0.0**	-	-
LBP+motion [21]	4.5	5.1	-	-
SBIQF+OFM [42]	0.83	**0.0**	5.8	-
Color LBP [16]**	1.5	5.1	8,8	10.8
Color SURF [18]	0.1	2.2	2.8	2.2
Color texture [17]	0.4	2.8	**2.1**	4.9
Proposed method	1.2	4.2	4.6	1.5

* from https://repository.iiitd.edu.in/jspui/handle/123456789/138
** the results are re-computed using the frame based scenario.

Table 10: Comparison against state of the art (cross-database)

Train on:	CASIA		Replay		MSU	
Test on:	Replay	MSU	CASIA	MSU	CASIA	Replay
Motion [11]	45.2	-	47,9	-	-	-
LBP [11]	45.9	-	57.6	-	-	-
LBP-TOP [11]	49.7	-	60.6	-	-	-
magLBP+magHOOF [45]*	50.1	-	47.1	-	-	-
LBP+HOOF [46]**	35.4	-	44.6	-	-	-
Spectral cubes [15]	34.4	-	50.0	-	-	-
CompRep [19]***	29.3	-	35.4	-	-	-
IDA [14]***	26.9	-	43.7	-	-	-
CNN [12]	48.5	-	45,5	-	-	-
CNN+eyeblink [13]	12.4	-	31.6	-	-	-
Deep dictionary [20]	22.8	-	27.4	-	-	-
Color LBP [16] ****	37.9	21.0	35.4	33.0	45.7	44.8
Color SURF [18]	26.9	19.1	**23.2**	**31.8**	**24.3**	29.7
Color texture [17]	30.3	20.4	37.7	34.1	46.0	33.9
Proposed method	**9.6**	**19.8**	39.2	33.3	29.7	**21.4**

* from https://repository.iiitd.edu.in/jspui/handle/123456789/138
** from [20].
*** from [13].
**** the results are re-computed using the frame based scenario.

protocol (i.e. combining all types of spoofing scenarios) without any deep analysis on the effect of different factors such as input sensor or presentation attack instrument (PAI) variation on the generalization capability. In this article, we showed that a method might be actually able to generalize well under some conditions even though its plain overall performance is far from satisfying.

For our in-depth analysis, we considered the color texture based face anti-spoofing approach, which has shown to provide state of the art performance in face spoofing detection and promising generalization abilities in our previous works. We performed extensive cross-database evaluation of seven color texture descriptors on three face anti-spoofing databases and focused on attack-specific analysis, namely display and print attacks, to gain better insights into the generalization performance of the different methods. Furthermore, we investigated the robustness of a color texture-based approach combining two complementary descriptors, each handling a specific type of attacks (print and display attacks). Our experiments revealed that the method is able to generalize extremely well against display attacks launched at short distance and moderate performance is achieved in the case of print attacks. We provided a thorough comparison against state of the art in both intra-database and inter-database scenarios, obtaining very promising results.

This work is by no mean complete. While excellent results were obtained in display attack detection, the color texture based approach might not be efficient in the case of print attacks. Thus, novel (complementary) approaches are needed for tackling the problem of print attacks. However, it is worth pointing out that the aim of this work was not to optimize the facial color texture representations for the different attack types but to demonstrate the limitations of the preliminary cross-database studies and to highlight the importance of careful breakdown analysis across different covariates. We believe and hope that our present work will advance the research and open new directions in face anti-spoofing.

Acknowledgements

The financial support from the Finnish Foundation for Technology Promotion and the Academy of Finland is acknowledged.

References

[1] S. Marcel, M. Nixon, S. Li (Eds.), Handbook of Biometric Anti-Spoofing, Springer-Verlag, 2014.
[2] B. Biggio, Z. Akhtar, G. Fumera, G. L. Marcialis, F. Roli, Security evaluation of biometric authentication systems under real spoofing attacks, IET Biometrics 1 (2012) 11–24.
[3] A. Hadid, N. Evans, S. Marcel, J. Fierrez, Biometrics systems under spoofing attack: An evaluation methodology and lessons learned, IEEE Signal Processing Magazine 32 (2015) 20–30.
[4] I. Chingovska, N. Erdogmus, A. Anjos, S. Marcel, Face recognition systems under spoofing attacks, in: T. Bourlai (Ed.), Face Recognition Across the Imaging Spectrum, Springer, 2016, pp. 165–194.
[5] Y. Li, Y. Li, K. Xu, Q. Yan, R. Deng, Empirical study of face authentication systems under osnfd attacks, IEEE Transactions on Dependable and Secure Computing (2016).
[6] ISO/IEC JTC 1/SC 37 Biometrics, Information technology – Biometric presentation attack detection – Part 1: Framework, International Organization for Standardization, 2016. Https://www.iso.org/obp/ui/#iso:std:iso-iec:30107:-1:ed-1:v1:en.
[7] A. Anjos, J. Komulainen, S. Marcel, A. Hadid, M. Pietikäinen, Face anti-spoofing: visual approach, in: S. Marcel, M. S. Nixon, S. Z. Li (Eds.), Handbook of biometric anti-spoofing, Springer, 2014, pp. 65–82.
[8] J. Galbally, S. Marcel, J. Fiérrez, Biometric antispoofing methods: A survey in face recognition, IEEE Access 2 (2014) 1530–1552.

[9] R. Ramachandra, C. Busch, Presentation attack detection methods for face recognition systems: A comprehensive survey, ACM Computing Surveys 50 (2017) 8:1–8:37.

[10] I. Chingovska, A. Anjos, S. Marcel, On the effectiveness of local binary patterns in face anti-spoofing, in: IEEE International Conference of the Biometrics Special Interest Group (BIOSIG), 2012.

[11] T. de Freitas Pereira, A. Anjos, J. De Martino, S. Marcel, Can face anti-spoofing countermeasures work in a real world scenario?, in: International Conference on Biometrics (ICB), 2013.

[12] J. Yang, Z. Lei, S. Z. Li, Learn convolutional neural network for face anti-spoofing, CoRR abs/1408.5601 (2014).

[13] K. Patel, H. Han, A. K. Jain, Cross-database face antispoofing with robust feature representation, in: Chinese Conference on Biometric Recognition (CCBR), 2016, pp. 611–619.

[14] D. Wen, H. Han, A. Jain, Face spoof detection with image distortion analysis, IEEE Transactions on Information Forensics and Security 10 (2015) 746–761.

[15] A. Pinto, H. Pedrini, W. Robson Schwartz, A. Rocha, Face spoofing detection through visual codebooks of spectral temporal cubes, IEEE Transactions on Image Processing 24 (2015) 4726–4740.

[16] Z. Boulkenafet, J. Komulainen, A. Hadid, Face anti-spoofing based on color texture analysis, in: International Conference on Image Processing (ICIP), 2015.

[17] Z. Boulkenafet, J. Komulainen, A. Hadid, Face spoofing detection using colour texture analysis, IEEE Transactions on Information Forensics and Security 11 (2016) 1818–1830.

[18] Z. Boulkenafet, J. Komulainen, A. Hadid, Face antispoofing using speeded-up robust features and fisher vector encoding, IEEE Signal Processing Letters 24 (2017) 141–145.

[19] K. Patel, H. Han, A. K. Jain, Secure face unlock: Spoof detection on smartphones, IEEE Transactions on Information Forensics and Security 11 (2016) 2268–2283.

[20] I. Manjani, S. Tariyal, M. Vatsa, R. Singh, A. Majumdar, Detecting silicone mask based presentation attack via deep dictionary learning, IEEE Transactions on Information Forensics and Security (2017).

[21] J. Komulainen, A. Anjos, A. Hadid, S. Marcel, M. Pietikäinen, Complementary countermeasures for detecting scenic face spoofing attacks, in: International Conference on Biometrics (ICB), 2013.

[22] N. Erdogmus, S. Marcel, Spoofing attacks to 2D face recognition systems with 3d masks, in: IEEE International Conference of the Biometrics Special Interest Group (BIOSIG), 2013.

[23] R. Raghavendra, K. B. Raja, C. Busch, Presentation attack detection for face recognition using light field camera, IEEE Transactions on Image Processing 24 (2015) 1060–1075.

[24] I. Pavlidis, P. Symosek, The imaging issue in an automatic face/disguise detection system, in: Proceedings of the IEEE Workshop on Computer Vision Beyond the Visible Spectrum: Methods and Applications (CVBVS), 2000, pp. 15–24.

[25] Z. Zhang, D. Yi, Z. Lei, S. Z. Li, Face liveness detection by learning multispectral reflectance distributions, in: IEEE International Conference on Automatic Face and Gesture Recognition (FG 2011), 2011, pp. 436–441.

[26] E. M. Rudd, M. Gnther, T. E. Boult, Paraph: Presentation attack rejection by analyzing polarization hypotheses, in: IEEE Conference on Computer Vision and Pattern Recognition Workshops (CVPRW), 2016, pp. 171–178.

[27] K. Kollreider, H. Fronthaler, M. I. Faraj, J. Bigun, Real-time face detection and motion analysis with application in liveness assessment, IEEE Transactions on Information Forensics and Security 2 (2007) 548–558.

[28] E. S. Ng, A. Y. S. Chia, Face verification using temporal affective cues, in: International Conference on Pattern Recognition (ICPR), 2012, pp. 1249–1252.

[29] G. Chetty, M. Wagner, Liveness verification in audio-video speaker authentication, in: Australian International Conference on Speech Science and Technology, 2004, pp. 358–363.

[30] R. W. Frischholz, A. Werner, Avoiding replay-attacks in a face recognition system using head-pose estimation, in: IEEE International Workshop on Analysis and Modeling of Faces and Gestures, 2003.

[31] M. De Marsico, M. Nappi, D. Riccio, J.-L. Dugelay, Moving face spoofing detection via 3D projective invariants, in: International Conference on Biometrics (ICB), 2012.

[32] T. Wang, J. Yang, Z. Lei, S. Liao, S. Z. Li, Face liveness detection using 3D structure recovered from a single camera, in: International Conference on Biometrics (ICB), 2013.

[33] J. Li, Y. Wang, T. Tan, A. K. Jain, Live face detection based on the analysis of fourier spectra, in: Biometric Technology for Human Identification, 2004, pp. 296–303.

[34] X. Tan, Y. Li, J. Liu, L. Jiang, Face liveness detection from a single image with sparse low rank bilinear discriminative model, in: European Conference on Computer Vision (ECCV), 2010, pp. 504–517.

[35] J. Bai, T.-T. Ng, X. Gao, Y.-Q. Shi, Is physics-based liveness detection truly possible with a single image?, in: IEEE International Symposium on Circuits and Systems (ISCAS), 2010, pp. 3425–3428.

[36] J. Määttä, A. Hadid, M. Pietikäinen, Face spoofing detection from single images using micro-texture analysis, in: International Joint Conference on Biometrics (IJCB), 2011.

[37] J. Yang, Z. Lei, S. Liao, S. Z. Li, Face liveness detection with component dependent descriptor, in: International Conference on Biometrics (ICB), 2013.

[38] N. Erdogmus, S. Marcel, Spoofing in 2D face recognition with 3D masks and anti-spoofing with kinect, in: Biometrics: Theory, Applications and Systems Conference (BTAS), 2013.

[39] N. Kose, J.-L. Dugelay, Countermeasure for the protection of face recognition systems against mask attacks, in: International Conference on Automatic Face and Gesture Recognition (FG 2013), 2013.

[40] J. Galbally, S. Marcel, Face anti-spoofing based on general image quality assessment, in: International Conference on Pattern Recognition (ICPR), 2014, pp. 1173–1178.

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

[42] L. Feng, L.-M. Po, Y. Li, X. Xu, F. Yuan, T. C.-H. Cheung, K.-W. Cheung, Integration of image quality and motion cues for face anti-spoofing: A neural network approach, Journal of Visual Communication and Image Representation 38 (2016) 451–460.

[43] G. Pan, Z. Wu, L. Sun, Liveness detection for face recognition, in: Recent Advances in Face Recognition, In-Teh, 2008, pp. 109–124.

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

[45] S. Bharadwaj, T. I. Dhamecha, M. Vatsa, S. Richa, Computationally efficient face spoofing detection with motion magnification, in: IEEE Conference on Computer Vision and Pattern Recognition Workshops (CVPRW), 2013.

[46] T. Siddiqui, S. Bharadwaj, T. Dhamecha, A. Agarwal, M. Vatsa, R. Singh, N. Ratha, Face anti-spoofing with multifeature videolet aggregation, in: International Conference on Pattern Recognition (ICPR), 2016.

[47] I. Chingovska, J. Yang, Z. Lei, D. Yi, S. Z. Li, O. Kähm, C. Glaser, N. Damer, A. Kuijper, A. Nouak, J. Komulainen, T. Pereira, S. Gupta, S. Khandelwal, S. Bansal, A. Rai, T. Krishna, D. Goyal, M.-A. Waris, H. Zhang, I. Ahmad, S. Kiranyaz, M. Gabbouj, R. Tronci, M. Pili, N. Sirena, F. Roli, J. Galbally, J. Fierrez, A. Pinto, H. Pedrini, W. S. Schwartz, A. Rocha, A. Anjos, S. Marcel1, The 2nd competition on counter measures to 2D face spoofing attacks, in: International Conference on Biometrics (ICB), 2013.

[48] X. Li, J. Komulainen, G. Zhao, P.-C. Yuen, M. Pietikäinen, Generalized face anti-spoofing by detecting pulse from face videos, in: International Conference on Pattern Recognition (ICPR), 2016.

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

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

[51] R. Nosaka, Y. Ohkawa, K. Fukui, Feature extraction based on co-occurrence of adjacent local binary patterns, in: Advances in Image and Video Technology, volume 7088 of *Lecture Notes in Computer Science*, Springer, 2012, pp. 82–91.

[52] R. Nosaka, C. H. Suryanto, K. Fukui, Rotation invariant co-occurrence among adjacent lbps, in: ACCV, Springer, 2012, pp. 15–25.

[53] V. Ojansivu, J. Heikkilä, Blur insensitive texture classification using local phase quantization, in: Image and Signal Processing, volume 5099 of

Lecture Notes in Computer Science, Springer, 2008, pp. 236–243.

[54] J. Kannala, E. Rahtu, Bsif: Binarized statistical image features, in: International Conference on Pattern Recognition (ICPR), 2012, pp. 1363–1366.

[55] H. Bay, T. Tuytelaars, L. Van Gool, Surf: Speeded up robust features, in: European Conference on Computer Vision (ECCV), Springer, 2006, pp. 404–417.

[56] F. Perronnin, J. Sanchez, T. Mensink, Improving the fisher kernel for large-scale image classification, in: European Conference on Computer Vision (ECCV), Springer, 2010, pp. 143–156.

[57] J. Sánchez, F. Perronnin, T. Mensink, J. Verbeek, Image classification with the fisher vector: Theory and practice, International Journal of Computer Vision 105 (2013) 222–245.

[58] Z. Zhang, J. Yan, S. Liu, Z. Lei, D. Yi, S. Z. Li, A face antispoofing database with diverse attacks, in: International Conference on Biometrics (ICB), 2012, pp. 26–31.

[59] Z. Zhang, J. Yan, S. Liu, Z. Lei, D. Yi, S. Li, A face antispoofing database with diverse attacks, in: International Conference on Biometrics (ICB), 2012, pp. 26–31.

[60] J. Komulainen, A. Hadid, M. Pietikäinen, Context based face antispoofing., in: International Conference on Biometrics: Theory, Applications and Systems (BTAS), 2013.

[61] M. Guillaumin, J. Verbeek, C. Schmid, Is that you? metric learning approaches for face identification, in: IEEE International Conference on Computer Vision (ICCV), 2009, pp. 498–505.

[62] A. Anjos, S. Marcel, Counter-measures to photo attacks in face recognition: a public database and a baseline, in: International Joint Conference on Biometrics (IJCB), 2011.

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