Face Spoofing Detection Using Colour Texture Analysis

Zinelabidine Boulkenafet, Jukka Komulainen, and Abdenour Hadid

***Abstract*— Research on non-intrusive software-based face spoofing detection schemes has been mainly focused on the analysis of the luminance information of the face images, hence discarding the chroma component, which can be very useful for discriminating fake faces from genuine ones. This paper introduces a novel and appealing approach for detecting face spoofing using a colour texture analysis. We exploit the joint colour-texture information from the luminance and the chrominance channels by extracting complementary low-level feature descriptions from different colour spaces. More specifically, the feature histograms are computed over each image band separately. Extensive experiments on the three most challenging benchmark data sets, namely, the CASIA face anti-spoofing database, the replay-attack database, and the MSU mobile face spoof database, showed excellent results compared with the state of the art. More importantly, unlike most of the methods proposed in the literature, our proposed approach is able to achieve stable performance across all the three benchmark data sets. The promising results of our cross-database evaluation suggest that the facial colour texture representation is more stable in unknown conditions compared with its gray-scale counterparts.**

***Index Terms*— Face recognition, spoofing detection, presentation attack, colour texture analysis.**

I. Introduction

NOWADAYS, it is known that most of existing face recognition systems are vulnerable to spoofing attacks. A spoofing attack occurs when someone tries to bypass a face biometric system by presenting a fake face in front of the camera. For instance, in [1], researchers inspected the threat of the online social networks based facial disclosure against the latest version of six commercial face authentication systems (Face Unlock, Facelock Pro, Visidon, Veriface, Luxand Blink and FastAccess). While on average only 39% of the images published on social networks can be successfully used for spoofing, the relatively small number of usable images was enough to fool face authentication software of 77% of the 74 users. Also, in a live demonstration during the International Conference on Biometric (ICB 2013), a female intruder with a specific make-up succeeded in fooling a face recognition system.[1] These two examples among many others highlight the vulnerability of face recognition systems to spoofing attacks.

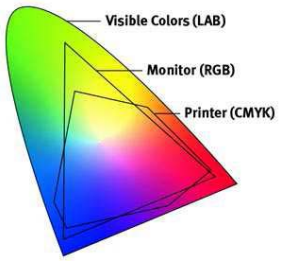


Fig. 1. Colour gamut of a monitor (RGB) and a printer (CMYK) compared with the whole pallet of visible colours [13].

Assuming that there are inherent disparities between genuine faces and artificial material that can be observed in single images (or a sequence of images), many anti-spoofing techniques analysing static (and dynamic) facial appearance properties have been proposed. The key idea is that an image of a fake face 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 fake face image is likely to have lower image quality compared to a genuine one captured in the same conditions due to e.g. lack of high frequency information [2]–[6]. Furthermore, the recaptured images may suffer from other quality issues, such as content-independent printing artefacts or video noise signatures [7]–[12]. In the literature, the facial appearance analysis based methods are usually referred to as texture or image quality analysis based techniques because the aforementioned properties can be considered as variations in the facial texture information or image quality.

The recapturing process described above introduces also inherent disparities in the colour information between a genuine face and a recaptured face image. This is due to the used spoofing medium (printed photograph, display device or mask) dependent gamut and other imperfections in the colour reproduction, e.g. printing defects or noise signatures. The camera used for capturing the targeted face sample will also lead to imperfect colour reproduction compared to the legitimate biometric sample. In general, printing and display devices have limited colour gamut compared to the whole pallet of visible colours (see, Figure 1). Moreover, images tend

Manuscript received August 27, 2015; revised January 7, 2016 and March 4, 2016; accepted March 29, 2016. Date of publication April 20, 2016; date of current version May 19, 2016. This work was supported by the Academy of Finland, Infotech Oulu, the Nokia Foundation, Northwestern Polytechnical University, and the Shaanxi Province. The associate editor coordinating the review of this manuscript and approving it for publication was Prof. Zhenan Sun.

Z. Boulkenafet and J. Komulainen are with the University of Oulu, FI-90014, Finland (e-mail: zboulken@ee.oulu.fi; jukmaatt@ee.oulu.fi).

A. Hadid is with Northwestern Polytechnical University, Xi'an 710129, China, and also with the University of Oulu, FI-90014, Finland (e-mail: hadid@ee.oulu.fi).

Color versions of one or more of the figures in this paper are available online at http://ieeexplore.ieee.org.

Digital Object Identifier 10.1109/TIFS.2016.2555286

[1] https://www.tabularasa-euproject.org/evaluations/tabula-rasa-spoofing-challenge-2013

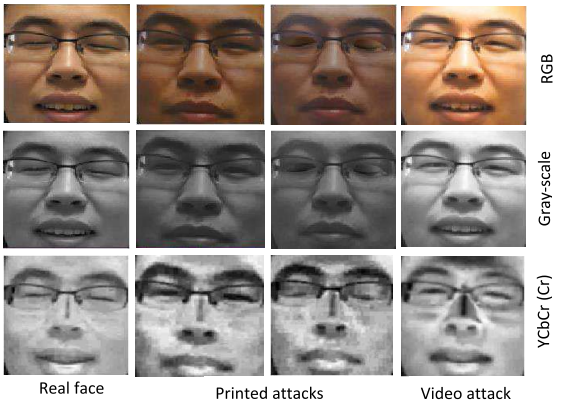


Fig. 2. Example of a genuine face and corresponding print and video attacks in RGB, gray-scale and YCbCr colour space.

to look different when they are printed or displayed using different devices. In order to preserve the colour and appearance perception across various devices, colour mapping algorithms can be applied on the source image to map its out-of-gamut colour into the colour gamut of a specific output device. However, these kinds of mapping functions can cause variation between the texture of the original and the output images.

Research on non-intrusive software-based face spoofing detection has mainly been focusing on analysing gray-scale images and hence discarding the colour information which can be a vital visual cue for discriminating fake faces from genuine ones. In a very recent work, Wen *et al.* [6] proposed colour quality based features that describe the chromatic degradation and the lack of colour diversity of recaptured face images. However, the actual local variations in colour texture information was not exploited for face spoofing detection.

Texture analysis of gray-scale face images can provide sufficient means to reveal the recapturing artefacts of fake faces if the image resolution (quality) is good enough to capture the fine details of the observed face. However, if we take a close look at the cropped facial images of a genuine human face and corresponding fake ones in Figure 2, it is basically impossible to explicitly name any textural differences between them because the input image resolution is not high enough.

To emulate the colour perception properties of the human visual system, colour mapping algorithms give a huge importance to the preservation of the spatially local luminance variations at the cost of the chroma information [14]. Human eye is indeed more sensitive to luminance than to chroma, thus fake faces still look very similar to the genuine ones when the same facial images are shown in colour (see, Figure 2). However, if only the corresponding chroma component is considered, some characteristic differences can be already noticed. While the gamut mapping and other artefacts cannot be observed clearly in the gray-scale or colour images, they are very distinctive in the chrominance channels. Thereby, colour texture analysis of the chroma images can be used for detecting these gamut mapping and other (colour) reproduction artefacts.

This present work extends our preliminary colour texture based approach presented in [15] and provides an in-depth analysis on the use of colour texture analysis for face spoofing detection. In addition to the colour local binary patterns (CLBP) descriptor [16] used in our prior work [15], we explore the facial colour texture content using four other descriptors: the local phase quantization (LPQ), the co-occurrence of adjacent local binary patterns (CoALBP), the binarized statistical image features (BSIF) and the scale-invariant descriptor (SID) that have already shown to be effective in gray-scale texture based face anti-spoofing [8], [17]. Here, we use these features for analysing colour texture by extracting face descriptions from different colour bands. To gain insight into which colour spaces are most suitable for discriminating genuine faces from fake ones, we considered three colour spaces, namely RGB, HSV and YCbCr in our experiments. The performance of the different facial colour texture representations is also compared to that of their gray-scale counterparts. Besides the CASIA Face Anti-Spoofing Database and the Replay-Attack Database, our proposed approach is also evaluated on the new MSU Mobile Face Spoof Database.

Among the significant contributions of this present work, we can cite:

- We provide a comprehensive review on the recent advances in face anti-spoofing.
- While most previous works on face spoofing detection are based on analysing only the luminance (i.e. gray-scale) information of the face images, we introduce a novel and appealing approach using colour texture analysis and demonstrate that the chroma component can be very useful in discriminating fake faces from genuine ones.
- We exploit the joint colour-texture information from the luminance and the chrominance channels by computing low-level features using different descriptors extracted from different colour spaces.
- We investigate how well different colour spaces and descriptors can be used for describing the intrinsic disparities in the colour texture between genuine faces and fake ones. We also perform a fusion study to analyze the complementarity of the different descriptors and different colour spaces.
- We conduct extensive experimental analysis on the three latest and most challenging spoofing databases using their pre-defined publicly well-defined experimental evaluation protocols ensuring the reproducibility of the results and a fair comparison against the state of the art.
- We provide an extensive comparative analysis against the state-of-the-art face spoofing detection techniques and show that our proposed approach outperforms all existing methods on two databases and achieves very competitive performance on the third database.
- Unlike most of the methods proposed in the literature, our proposed approach is able to achieve stable performance across all the three benchmark datasets. Furthermore, in our inter-database evaluation, the facial colour texture representation showed promising generalization capabilities, thus suggesting that colour texture can be more stable in unknown conditions compared to its gray-scale counterparts.

The rest of this article is organized as follows. Section II gives a comprehensive review on the recent advances

in face anti-spoofing. Section III describes our colour texture based countermeasure. Section IV presents our experimental data and setup. The experimental results are then introduced and discussed in Section V. Finally, Section VI draws concluding remarks and points out possible future research directions.

II. Prior Work

There exists no unified taxonomy for the different spoof detection approaches. In this artcile, a three-part categorization is followed dividing the individual face spoofing detection schemes into hardware-based, challenge-response and software-based techniques.

Hardware-based solutions using 3D [18] or multi-spectral [19], [20] imaging provide efficient means for detecting face spoofs because they offer additional useful information on the surface reflectance properties or depth of the observed face. For instance, a low-cost depth sensor, i.e. Microsoft Kinect, can be utilized for differentiating a real face from a planar surface, e.g. video display or photograph, in a quite straightforward manner [18]. Skin reflectance measurements at two specific wavelengths can be used to distinguish a genuine face from artificial materials used in 3D masks and 2D surfaces because human skin has extremely low reflectance in the upper-band of near-infrared (NIR) spectrum which is a universal property among human race [19], [20]. Thermal information can also be used for detecting prints and replayed videos. Adding and subtracting skin tissue using redistributed fat or creating or removing scars with silicone are typical operations of plastic surgery. Furthermore, surgical operations usually cause alteration in blood vessel flow that can be seen as cold spots in the thermal domain. These kinds of physiological changes can be detected in the thermal infrared (IR) region [19].

On the other hand, depth sensors are powerless under 3D mask attacks if depth cue is the only utilized countermeasure. It is a known fact that thermal radiation can pass through materials, which causes problems when thermal IR information is used against wearable mask attacks [20]. The existing NIR based techniques have also difficulties in capturing the reflectance disparities between genuine faces and 3D masks due to the 3D shape and variety of artificial materials [20]. Furthermore, the use of NIR imaging is restricted to indoor use only since the sunlight causes severe perturbation. The dedicated imaging solutions are indeed effective in detecting various kinds of artificial faces if they are coupled in the same system [19]. Unfortunately, the problem with hardware-based techniques is that, in general, they are either quite intrusive, expensive or impractical because unconventional imaging devices (e.g. active lighting) are required. Sensor-based techniques have been usually evaluated mainly to demonstrate a proof of concept or have not been experimentally validated at all in the worst case, like in [19]. Therefore, it is extremely hard to directly compare hardware-based approaches with other related biometric solutions.

It is worth mentioning, however, that multi-modal and hardware-based solutions are still worth considering. While nowadays every mobile phone and laptop are equipped with a microphone and camera, other sensors, such as 3D and NIR imaging, are emerging in mobile devices which opens up new possibilities for face anti-spoofing. Furthermore, the existing (mobile) devices already provide means for novel spoofing detection schemes. For instance, Smith *et al.* [21] proposed to analyse dynamic reflections from the observed person's face caused by varying illumination due to a sequence of images (digital watermarks) presented on the used display device, e.g. a tablet or a laptop, for validating that the biometric data was captured in real-time and not injected to the communication system channels (replay-attack detection). However, it would be probably possible to couple similar digital watermarks for performing both replay-attack and spoofing detection simultaneously.

User collaboration can also be used for revealing spoofing attacks because we humans tend to be interactive, whereas a photo or video replay attack cannot respond to randomly specified action requirements. In particular, a face authentication system prompts a user for a specific action (challenge), such as a facial expression [22], [23], mouth movement [22], [24] or head rotation (3D information) [25]–[27], and then analyses the user activity in order to check whether the required action was actually performed (response) or not.

The drawback of the challenge-response approach is that it requires user cooperation, thus making the authentication process a time-consuming and unpleasant experience. Another disadvantage of challenge-response based countermeasures is that it is rather easy to deduce which liveness cues need to be fooled. For instance, the request for uttering words suggests that analysis of synchronized lip movement and lip reading is utilized, whereas rotating head in a certain direction reveals that the 3D geometry of the head is measured. For non-intrusive approaches, it is usually not known which countermeasures are used, thus the system might be harder to deceive [28].

As a consequence, it would be rather appealing to utilize anti-spoofing techniques requiring basically no user-cooperation and using the conventional cameras included in the existing face authentication systems. Another advantage of non-intrusive software-based countermeasures is that they can be assessed on common benchmark datasets or, even better, if any new data is collected, it can be distributed to other researchers. Although the number of publicly available datasets is still quite scarce, new anti-spoofing databases appear gradually due to the increasing interest in anti-spoofing by the research community [3], [4], [6], [29], [30] and international competitions [31], [32]. The benchmark datasets have been indispensable tools for the researchers by providing them the opportunity to concentrate on investigating the problem of anti-spoofing. This has had a significant impact on the amount of papers on data-driven countermeasures during the recent years.

Non-intrusive software-based countermeasures can be categorized into static and dynamic techniques based on whether temporal information or features are utilized [33]. The dynamic methods in the related literature are mainly based on analysing the motion or liveness while the static methods are focused on analysing the facial appearance or quality

based cues. Therefore, the following taxonomy for non-intrusive software-based face spoofing detection schemes is based on the inspected visual cues: motion, facial appearance and context.

Typical non-intrusive software-based countermeasure to face spoofing is liveness detection that aims at detecting physiological signs of life, such as eye blinking [34]–[36], facial expression changes [22], [35], [36] mouth movements [22], [35], [36] and even small colour and motion changes that appear on the face due to the natural human blood flow [32]. In addition to facial motion used in liveness detection, other motion cues can also be exploited for face anti-spoofing. For example, it can be assumed that the movement of planar objects, e.g. video displays and photographs, differs significantly from real human faces which are complex non-rigid 3D objects [37], [38]. If a face spoof is not tightly cropped around the targeted face or it has an incorporated background scene (scenic fake face), it should be possible to observe high correlation between the overall motion of the face and the background regions for stationary face recognition systems [39], [40].

The main problem of liveness detection and motion analysis based anti-spoofing techniques is that the verification process takes some time or the user needs to be still very cooperative in practice. Even though motion is an important visual cue, vitality and non-rigid motion detectors relying only on spontaneous facial movements are powerless under video replay attacks. The lack of motion may lead to a high number of authentication failures if user cooperation (e.g. challenge-response) is not requested.

Assuming that the inherent disparities between genuine faces and artificial material can be observed in single images (or a sequence of images), another category of non-intrusive software-based anti-spoofing techniques is based on the analysis of static (and dynamic) facial appearance properties, such as reflectance, shading, texture and quality. Intuitively, the main advantage of single image based spoofing detection schemes is that they treat video playback attacks as if they were photo attacks, since individual video frames are considered [7].

One can assume that fake faces are usually smaller in size or they would contain fewer high frequency components compared to genuine ones, thus countermeasures based on analysing the high frequency content have been proposed [2]–[4]. Such an approach may work well for down-sampled photos or crude face masks but is likely to fail for higher-quality spoof samples.

Alternatively, it is likely that real faces and fake ones present different texture patterns because of facial texture quality degradation due to recapturing process and disparities in surface and reflectance properties. Hence, micro-texture analysis has been utilized for capturing these differences [7], [8]. Texture based face anti-spoofing has been widely adopted in face anti-spoofing research. Methods performing joint analysis of texture and local gradient structures have also been proposed [9], [10]. Preliminary studies in 3D mask attack detection [30], [41] besides print and video-replay attacks have also been reported. In addition to analysing the structure of facial textures, spatiotemporal texture analysis is applied for describing specific dynamic events, e.g. facial motion patterns and sudden characteristic reflections of planar spoofing media [11] and content-independent video noise signatures [12].

The major drawback of texture analysis based methods is that high resolution input images are required in order to extract the fine details needed for discriminating genuine faces from spoofing media. While lower imaging quality can be enough for detecting the most crude attack attempts, such as small mobile phone displays and prints with strong artefacts, the grid structure of a display device or facial pores can be captured only in high-definition close-up images. On the other hand, high false rejection rates can also be an issue if the acquisition quality is not good enough. Furthermore, the nature of texture patterns varies a lot due to different acquisition conditions and spoofing media, thus diverse datasets are probably needed for training the facial texture models, especially at conventional webcam image quality. The generalization capabilities of the texture based methods are not yet clear due to the lack of variation between training and test set, e.g. illumination, sensor quality and user demographics, let alone unknown attack scenarios. The initial inter-database tests [27], [42] have suggested that the performance of texture based techniques degrades dramatically when the face models learned from one dataset are tested on another dataset.

Pure image quality assessment based features have shown comparable performance to texture analysis based algorithms [5]. However, as expected, they are robust in detecting fake faces presented on mobile phone displays, whereas high-definition face spoofs caused problems [5]. In addition, the spoofing detection performance of the proposed feature set was highly dependent on the used imaging quality, i.e. the method performed well on high quality input images, whereas the results degraded dramatically at lower acquisition qualities [5].

In a very recent work, Wen *et al.* [6] argued that commonly used features, e.g. LBP, may be too person-specific or contain too much redundant information for face anti-spoofing because they are capable of capturing the facial details, i.e. differentiating individuals for face recognition purposes. Hence, they proposed to extract features that do not try to capture the facial details but the characteristic differences between genuine faces and fake ones, including characteristic reflection and quality properties, e.g. blur and colour diversity. The experimental validation showed promising generalization capabilities compared to texture based methods but only with short distance spoofing attacks. The features did also generalize to cameras with similar quality but not to cameras with distinctively different quality. However, their argument on features describing facial details suggests that person-specific anti-spoofing models [43], [44] can improve the generalization of texture based approaches, for instance.

In addition to motion [39], [40], the observed scene provides also actual contextual cues which have shown to be useful for anti-spoofing. For instance, scene context matching could be utilized for checking if the background scene of the stationary face recognition system suddenly changes [28]. Furthermore, a bezel (frame) of a display device or photograph edges, or

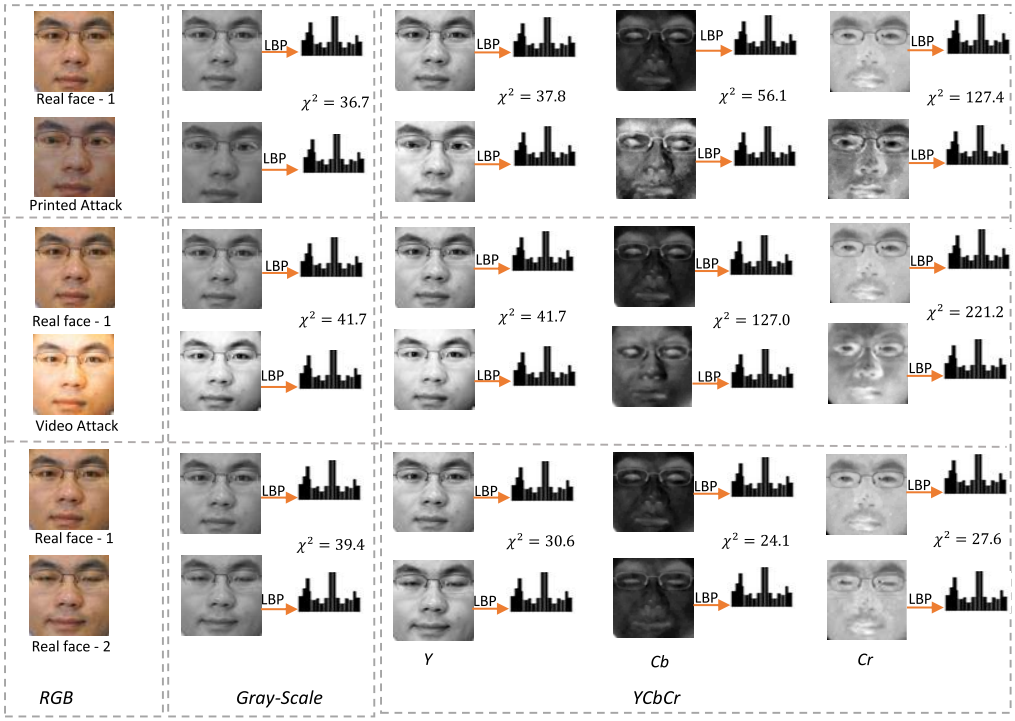


Fig. 3. The similarity between the holistic LBP descriptions extracted from genuine faces and fake ones. The original RGB images are shown on the left. In the middle, the similarity between the LBP descriptions extracted from the gray-scale images is presented. The similarity between the LBP descriptions extracted from the different colour channels in the YCbCr colour space is presented on the right.

the attackers hands might be visible in the provided view [45]. These kinds of contextual cues are rather straightforward to detect but they are also easy to conceal or cannot be exploited in certain use-case scenarios.

It is reasonable to assume that there is no universal anti-spoofing technique for detecting all kinds of attacks because every countermeasure most likely has its own vulnerability ("a golden fake") that can be exploited by an attacker. Therefore, system design is a very important research topic of its own. For instance, the robustness of generic anti-spoofing methods, i.e. treating all sorts of attack scenarios equally, could be improved by training an ensemble of separate detectors for the different display media, e.g. face prints, video displays and masks, like performed in [6]. Moreover, intuitively, a combination of several complementary countermeasures increases the robustness to various types of spoofing attacks [46]. Therefore, it is not surprising that fusion of several methods analysing the motion and facial appearance has been a common trend in the recently organized two competitions on software-based anti-spoofing [31], [32]. For instance, in the 2nd competition on counter measures to 2D face spoofing attacks [32], all the best-performing methods were utilizing some sort of combination of both motion and texture analysis. Other important directions for system-level research is the utilization of person-specific information either by investigating the joint operation of spoofing detection and actual face recognition stages [47] or person-specific anti-spoofing models [43], [44] that can improve both robustness and generalization capabilities of the existing software-based countermeasures.

III. Colour Texture Analysis Based Face Anti-Spoofing

Face spoofing attacks are most likely performed by displaying the targeted faces using prints, video displays or masks to the input sensor. The most crude attack attempts performed, e.g. using small mobile phone displays or prints with strong artifacts, can be detected by analysing the texture and the quality of the captured gray-scale face images. However, as shown in Figure 3, it is reasonable to assume that fake faces of higher quality are harder or nearly impossible to detect using only luminance information of webcam-quality images. In Figure 3, this effect is demonstrated by measuring the similarity between the local binary pattern (LBP) descriptions extracted from a genuine face (Real face 1), another genuine face (Real face 2) and two fake face images (Printed Attack and Video Attack) of the same person. The similarity is measured using the Chi-square distance:

$$d_{\chi 2}(H_x, H_y) = \sum_{i=1}^{N} \frac{(H_x(i) - H_y(i))^2}{H_x(i) + H_y(i)}, \tag{1}$$

where H_x and H_y are two LBP histograms with N bins. In addition to its simplicity, the Chi-square distance is shown to be effective to measure the similarity between two LBP histograms [48]. From Figure 3, we can observe that the Chi-square distance between gray-scale LBP histograms of the genuine face and the printed fake face is smaller than the one between two genuine face images. Moreover, the difference in similarity between the texture descriptions of genuine faces

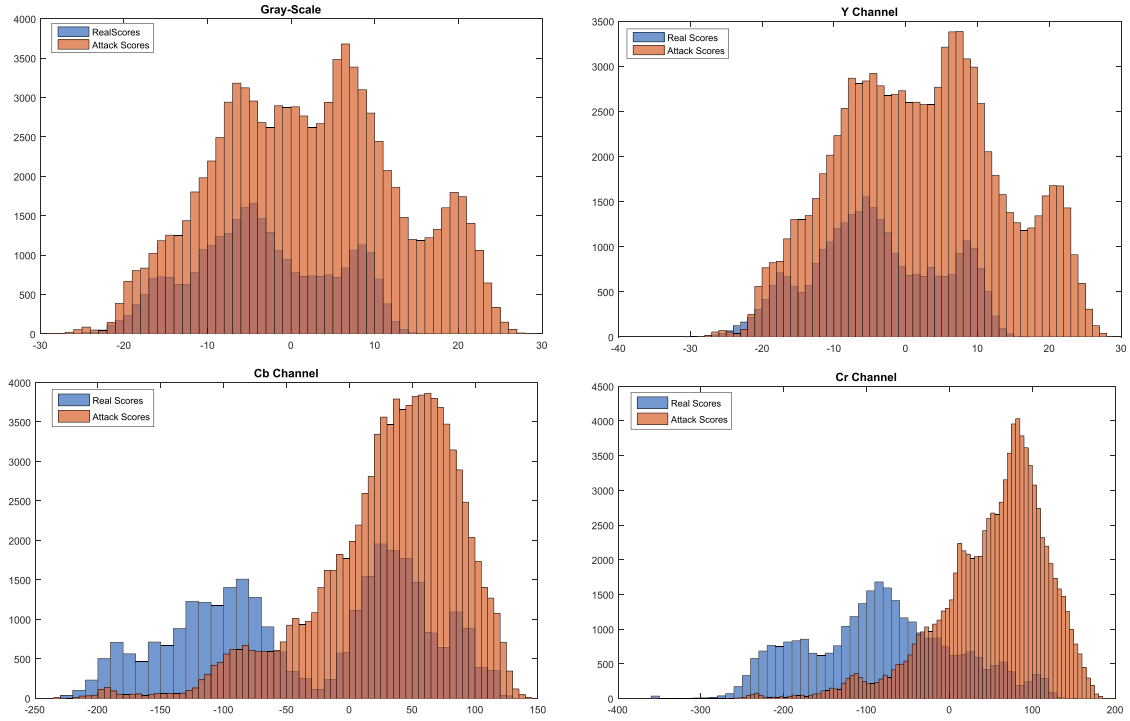


Fig. 4. The distribution of the real and attack scores on the Replay-Attack Database (test set) using simple Chi-square statistics on LBP descriptions extracted from gray-scale images and the individual channels of the YCbCr colour space.

and the Chi-square distance between the genuine face and the video attack is not significant. It is worth noting, however, that similarity measured with pure Chi-square distance does not necessarily indicate that there are no intrinsic disparities in the gray-scale texture representation that could be exploited for face spoofing detection.

Fortunately, the colour reproduction (gamut) of different display media, e.g. photographs, video displays and masks, is limited compared to genuine faces. Thus, the presented fake faces suffer from spoofing medium dependent colour. In addition, gamut mapping functions are typically required in order to preserve colour perception properties across different output devices, e.g. printer or video display, which can alter the (colour) texture of the original image. In general, the gamut mapping algorithms focus on preserving the spatially local luminance variations in the original images at the cost of the chrominance information because the human eye is more sensitive to luminance than to chroma [14]. Therefore, we humans cannot observe the evident differences when only the texture of the luminance information between the original and the transformed images is analysed. The camera used for capturing the targeted face sample will also lead to imperfect colour reproduction compared to the legitimate sample. Furthermore, a recaptured face image is likely to contain local and overall variations of colour due to other imperfections in the reproduction process of the targeted face. It is also worth mentioning that other disparities in facial texture, including printing defects, video artefacts, noise signatures of display devices and moiré effects, should be more evident in the original colour images compared to gray-scale images.

Both the display medium dependent colour signatures, including gamut mapping artefacts, and other intrinsic local variations in (chroma) texture due to the recapturing process (noise) can be described by analysing the colour texture of the chroma channels. As can be seen in Figure 3, the texture information of the chrominance components in YCbCr colour space show apparent disparities between the genuine faces and fake ones. The dissimilarity of the corresponding LBP descriptions is also significant, while the similarity between the descriptions of genuine faces still remains. Since the chrominance components are separated from the luminance information, they are also more tolerant to illumination variation assuming that the acquisition conditions are reasonable.

To confirm the observations in Figure 3, we have also conducted a statistical analysis on the Replay-Attack Database [29]. More specifically, we computed mean LBP histograms for both real and fake face images in the training set and used these two models to compute a Chi-square distance based score value for each sample in the test set as follows:

$$d(H_x, H_r, H_f) = d_{\chi 2}(H_x, H_r) - d_{\chi 2}(H_x, H_f), \qquad (2)$$

where H_x is the LBP histogram of the test sample, and H_r and H_f are the reference histograms for real and fake faces, respectively. Figure 4 illustrates the score distributions of the real faces and spoofs in the gray-scale space and in the three channels of the YCbCr colour space. The results confirm our hypothesis in the sense that the Chi-square statistics of the real and fake face descriptions in the gray-scale space and

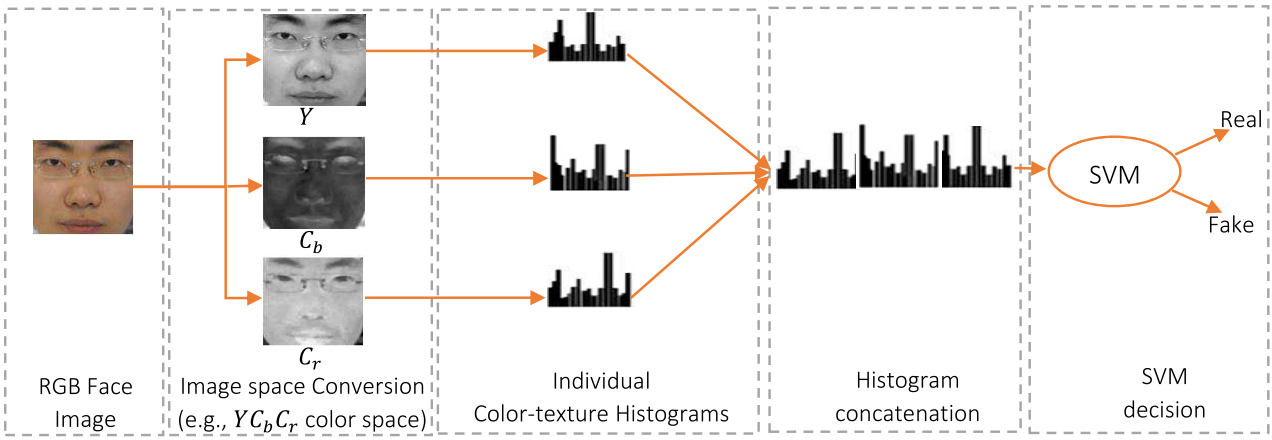


Fig. 5. Illustration of the proposed face anti-spoofing approach.

Y channel are overlapping while they are better separated in the chroma components of the YCbCr space.

In this present work, we aim to investigate the effectiveness of different texture descriptors more closely in detecting various kinds of face spoofs by extracting holistic face representations from luminance and chrominance images in different colour spaces. The general block diagram of the proposed face spoofing detection approach is depicted in Figure 5. First, the face is detected, cropped and normalised into an M×N pixel image. Then, holistic texture descriptions are extracted from each colour channel and the resulting feature vectors are concatenated into an enhanced feature vector in order to get an overall representation of the facial colour texture. The final feature vector is fed to a binary classifier and the output score value describes whether there is a live person or a fake one in front of the camera.

The facial representations extracted from different colour spaces using different texture descriptors can also be concatenated in order to benefit from their complementarity. This kind of fusion is investigated more closely in Section V-B. The proposed method can operate either on a single video frame or video sequences, thus practically real-time response can be achieved.

A. Colour Spaces

RGB is the most used colour space for sensing, representating and displaying colour images. However, its application in image analysis is quite limited due to the high correlation between the three colour components (red, green and blue) and the imperfect separation of the luminance and chrominance information. On the other hand, the different colour channels can be more discriminative for detecting recapturing artefacts, i.e. providing higher contrast for different visual cues from natural skin tones.

In this work, we considered two other colour spaces, HSV and YCbCr, to explore the colour texture information in addition to RGB. Both of these colour spaces are based on the separation of the luminance and the chrominance components. In the HSV colour space, hue and saturation dimensions define the chrominance of the image while the value dimension corresponds to the luminance. The YCbCr space separates the RGB components into luminance (Y), chrominance blue (Cb) and chrominance red (Cr). It is worth noting that the representation of chroma components in HSV and YCbCr spaces is different, thus they can provide complementary facial colour texture descriptions for spoofing detection. More details about these colour spaces can be found e.g. in [49].

B. Texture Descriptors

In principle, texture descriptors originally designed for gray-scale images can be applied on colour images by combining the features extracted from different colour channels. In this present study, the colour texture of the face images is analysed using five descriptors: Local Binary Patterns (LBP), Co-occurrence of Adjacent Local Binary Patterns (CoALBP), Local Phase Quantization (LPQ), Binarized Statistical Image Features (BSIF) and Scale-Invariant Descriptor (SID) that have shown to be very promising features in prior studies [8], [17] related to gray-scale texture based face anti-spoofing. Detailed descriptions of each of these features are presented in the following.

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

$$LBP_{P,R}(x, y) = \sum_{n=1}^{P} \delta(r_n - r_c) \times 2^{n-1}, \tag{3}$$

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 neighbourhood pixels located at the circle of radius R $(R > 0)$, respectively. The occurrences of the different binary patterns are collected into histogram to represent the image texture information. LBP pattern is defined as uniform if its binary code contains at most two transitions from 0 to 1 or from 1 to 0. For example 01110000 (2 transitions) and 00000000 (0 transitions) are uniform patterns.

2) Co-Occurrence of Adjacent Local Binary Patterns (CoALBP): In the original LBP descriptor, the packing of the LBP patterns into histogram tends to discard the spatial information between the patterns. To exploit the spatial relation between the patterns, the authors in [51] proposed the co-occurrence of adjacent local binary patterns method [51]. In this method, they first extract the LBP patterns from the images using simplified LBP descriptors (LBP_+ or $LBP_\times$), then to capture the correlation between the spatially adjacent patterns, four directions were defined: $D = \{(0, \Delta B), (\Delta B, 0), (\Delta B, \Delta B)$ and $(-\Delta B, \Delta B)\}$ where B is the distance between two adjacent LBP patterns. For each direction $d \in D$ a 16×16 2-D histogram is created. The resulting histograms are then reshaped and concatenated to form a final feature vector.

3) Local Phase Quantization (LPQ): The LPQ descriptor [52] was proposed to deal with the blurred images. It uses the local phase information extracted by a Short Term Fourier Transform (STFT) to analysis the $M \times M$ neighbourhoods 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 considered: $u_0 = (\alpha, 0)$, $u_1 = (\alpha, \alpha)$, $u_2 = (0, \alpha)$, $u_3 = (-\alpha, -\alpha)$ where α is a small scalar frequency ($\alpha << 1$). These frequencies correspond to the directions 0, 45, 90 and 135, respectively. The basic LPQ features at the pixel position x are represented by a 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 part and the imaginary part of a complex number. The element of this vector are then quantized using the δ function defined previously. Finally, the resulting binary quantized coefficients are represented as integer value in [0-255] and collected into histogram. To make the LPQ coefficients statistically independents a de-correlation step based on the whitening transform can be applied before the quantization process.

4) Binarized Statistical Image Features (BSIF): Similarly to LBP, the BSIF descriptor [53] computes a binary code string for each pixel in an image where each bit is obtained by first convolving the image with a linear filter and then binarizing the filter responses. The number of the used filters determines the length of the binary code. In order to obtain statistically meaningful representation of the image data and efficient encoding using simple element-wise quantization, the fixed set linear filters are learnt from a set of image patches by maximizing the statistical independence of the filter responses using independent component analysis (ICA). In our experiments, we used the set of filters provided by the authors of [53] that were learnt from a set of natural image patches.

5) Scale-Invariant Descriptor (SID): The SID feature is based on the shift property of Fourier transform, i.e. its magnitude is invariant to translations. To be more specific, if an image is first re-sampled densely enough on a log-polar grid, rotations and scalings in the original image domain are equivalent to translations on the new sampling grid. Thus, when Fourier transform is applied on the re-sampled image, invariance to both scale and rotation is achieved (but at the cost of high dimensionality due to dense sampling).

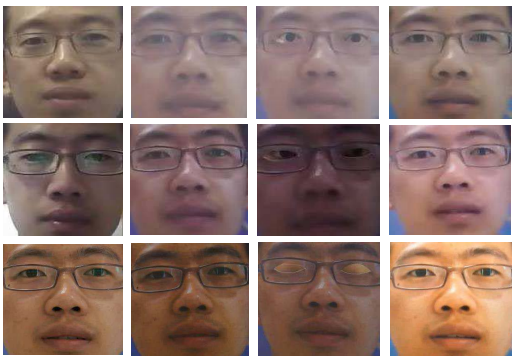


Fig. 6. Cropped and normalized example face images from the CASIA FASD. From top to bottom: low , normal and high quality images. From the left to the right: real faces and the corresponding warped photo, cut photo and video replay attacks.

To summarize, let I be a face image represented in a colour space S and, let $H_s^{(i)}, \{i = 1 : L\}$ be its texture histograms extracted from the L channels of the space S. The colour texture features of the image I represented in the space S can be defined by:

$$H_s = [H_s^{(1)} \dots H_s^{(L)}]. \quad (4)$$

IV. Benchmark Datasets and Experimental Setup

To assess the effectiveness of our proposed anti-spoofing technique, we considered three latest face anti-spoofing databases: CASIA Face Anti-Spoofing Database (CASIA FASD), Replay-Attack Database and MSU Mobile Face Spoof Database (MSU MFSD). These three datasets are the most challenging face anti-spoofing benchmark databases that consist of recordings of real client accesses and various spoofing attack attempts captured with different imaging qualities, including mobile phones, webcams and digital system cameras. In the following, a brief description of the three databases is given.

A. CASIA Face-Anti-Spoofing Database (CASIA FASD)

The CASIA Face Anti-Spoofing Database [4] contains video recordings of genuine and fake faces (see, Figure 6, for cropped and normalized example images). 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 attack 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).

B. Replay-Attack Database

The Replay-Attack Database [29] consists of video recordings of real accesses and attack attempts to 50 clients

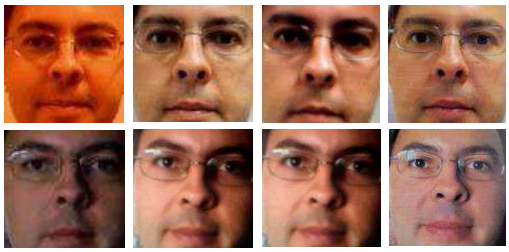


Fig. 7. Cropped and normalized example face images from the Replay-Attack Database. The first row presents images taken from the controlled scenario, while the second row corresponds to the images from the adverse scenario. From the left to the right: real faces and the corresponding high definition, mobile and print attacks.

Fig. 8. Cropped and normalized example face images from the MSU MFSD. The first row corresponds to images that have been taken with an Android phone, while the second row shows images captured with a laptop camera. From the left to the right: real faces and the corresponding iPad, iPhone and print attacks.

(see, Figure 7, for cropped and normalized example images). 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. According to the support used in presenting the fake face devices in front of the camera, two types of attacks were defined: *hand based attacks*, i.e. the attack devices were held by the operator and *fixed-support attacks*, i.e. the attack devices were set on a fixed support. For the evaluation, the 50 subjects were divided on three subject-disjoint subsets for training, development and testing.

C. MSU Mobile Face Spoof Database (MSU MFSD)

The MSU Mobile Face Spoof Database [6] consists of 280 video recordings of real and fake faces (see, Figure 8, for cropped and normalized example images). 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 laptop camera, the videos were taken with a resolution of 640×480, while for the Android camera, the videos were captured with a resolution of 720×480. The duration of each video is at least nine seconds.

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 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 (5184×3456) of the subject's faces were taken with the Canon 550D camera which then were printed on A3 paper using an HP colour laserjet CP6015xh printer. For the performance evaluation, the 35 subjects of the MSU MFSD database were divided into two subject-disjoint subsets for the training and testing (15 and 20 subjects, respectively).

D. Experimental Setup

In our experiments, we followed the official overall test protocols of the three databases which allows a fair comparison with other methods proposed in the literature. Since the CASIA FASD and MSU MFSD lack a pre-defined validation set, 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 prior works, including [3], [5], [8], [11], [29], [43], and [46], and to mitigate the effect of face normalization, all texture descriptions were extracted from face images that were aligned based on eye locations and normalized into 64×64 pixels with interpupillary distance of 32 pixels. The eyes were localized using the PittPatt 5.2.2 SDK [54]. For the LBP descriptor, we used the $\text{LBP}_{8,1}$ operator (i.e. $P = 8$ and $R = 1$) to extract the micro texture patterns. In addition to the basic LBP descriptor (LBP), we have also evaluated the LBP with uniform mapping (LBP^{u2}). In the case of the CoALBP descriptor, the features were a concatenation of three histograms computed using the LBP_{+} operator with radius $R = \{1, 2, 4\}$ and the corresponding directions defined by the distances $B = \{2, 4, 8\}$. The parameters for the LPQ operator were: $M = 7$, $\alpha = 1/7$ and $\sigma = 0.9$. BSIF features are obtained using eight filters of size 7×7. For the SID descriptor the parameters were: the number of rays $= 32$, the number of rings $= 28$, the number of derivative orientations $= 3$, the maximum ring radius $= 231$,

TABLE I

THE PERFORMANCE ON THE DEVELOPMENT AND TEST SET OF THE REPLAY-ATTACK DATABASE IN TERMS OF EER (%) AND HTER (%), RESPECTIVELY

Method	Gray		RGB		HSV		YCbCr	
	EER	HTER	EER	HTER	EER	HTER	EER	HTER
LBP^{u2}	17.9	13.7	4.6	6.8	6.9	10.6	2.3	5.6
CoALBP	12.9	16.7	6.2	8.0	3.7	4.3	**1.4**	**4.7**
LPQ	25.3	31.1	9.7	10.3	7.9	9.2	6.3	11.5
BSIF	31.5	30.8	13.5	11.3	8.2	10.3	10.9	10.7
SID	22.2	21.8	14.5	12.3	3.0	8.7	4.9	11.2

and the minimum ring radius = 3. More details about these parameters can be found in the original papers.

To capture both the texture variations of facial appearance and dynamics, we average the feature descriptions within a time window of three seconds. The whole video length of the provided training sequences is exploited for learning the colour texture models. More specifically, time windows with an overlap of two seconds are utilized in order to get more training data. In the test stage, only the average of the features within the first three seconds is used to classify each video. In Section V-C, the per frame spoofing detection performance is also reported for the proposed final facial colour texture representation. The binary classification was performed using a linear support vector machine (SVM) classifier (using LIBLINEAR [55]).

V. EXPERIMENTAL RESULTS AND ANALYSIS

In this section, we present and discuss the results obtained using the different colour texture descriptors on the different colour spaces. We begin our experiments by comparing the performances of the colour texture features and their gray-scale counterparts. Then, we will combine complementary facial colour texture representations to form the final face description used in our anti-spoofing method and compare its performance against the state-of-the-art algorithms. Finally, we evaluate the generalization capabilities of the proposed approach by conducting cross-database experiments.

A. Effectiveness of the Colour Texture Analysis

In this part, we present the performance of the different feature descriptors extracted from the different image representations. Tables I, II and III present the performance of the different feature descriptors applied on the gray-scale, RGB, HSV and YCbCr image representations. It can be clearly seen that the utilization of colour texture information significantly improves the robustness of different descriptors compared to their gray-scale counterparts. When it comes to the different colour spaces, we observe that the use of YCbCr and HSV colour spaces yields generally in better performance compared to RGB colour space. On CASIA FASD and Replay-Attack Database, the CoALBP features extracted from the HSV and the YCbCr colour spaces provide the best results. Compared with the gray-scale counterparts, the CoALBP colour-texture features improve the performance on CASIA and Replay-Attack databases with relative percentages of 63.5% and 71.6%, respectively. On MSU MFSD, the best results are obtained using the LPQ descriptor where the features extracted from the YCbCr colour space improve the performance with a relative percentage of 69.0% compared with the gray-scale LPQ features.

TABLE II

THE PERFORMANCE ON THE TEST SET OF THE CASIA FASD REPORTED IN TERMS OF EER (%)

Method	Gray	RGB	HSV	YCbCr
LBP^{u2}	22.6	21.0	13.6	12.4
CoALBP	14.8	11.0	**5.5**	10.0
LPQ	23.2	14.4	7.4	16.2
BSIF	26.2	21.0	6.7	17.0
SID	19.9	15.8	11.2	11.6

TABLE III

THE PERFORMANCE ON THE TEST SET OF THE MSU MFSD REPORTED IN TERMS OF EER (%)

Method	Gray	RGB	HSV	YCbCr
LBP^{u2}	35.0	12.3	13.9	13.0
CoALBP	19.9	17.7	9.8	8.1
LPQ	23.9	23.2	12.2	**7.4**
BSIF	24.2	23.5	12.1	7.5
SID	24.4	22.3	13.5	8.5

TABLE IV

THE FUSION PERFORMANCES OF THE COALBP AND THE LPQ DESCRIPTORS ON THE CASIA FASD, MSU MFSD AND REPLAY-ATTACK

Colour space	CASIA FASD	MSU MFSD	Replay-Attack	
	EER	EER	EER	HTER
RGB	8.6	12.5	9.6	10.3
HSV	4.0	11.4	3.7	4.3
YCbCr	8.7	4.9	0.3	4.6
HSV+RGB	3.6	8.7	6.1	7.6
YCbCr+RGB	4.4	4.9	2.5	7.2
HSV+YCbCr	**3.2**	**3.5**	**0.0**	**3.3**

B. Fusion of Complementary Colour Texture Representations

As can be seen in the previous experiments, the different facial colour texture representations, including colour spaces and feature descriptors, seem to perform better on different datasets. In order to benefit from the potential complementarity of the CoALBP and the LPQ face descriptions, we propose to fuse them by concatenating their resulting histograms. The length of the final feature vector is 37788. Table IV depicts that the combined facial texture representations are indeed complementary improving the performance of the individual descriptors on all the three databases.

To further enhance the facial colour texture representation, the features extracted from the different colour spaces are also

TABLE V

COMPARISON BETWEEN THE PROPOSED COUNTERMEASURE AND STATE-OF-THE-ART METHODS ON THE THREE BENCHMARK DATASETS USING THE VIDEO BASED EVALUATION

Method	Replay-Attack		CASIA	MSU
	EER	HTER	EER	EER
DoG [4]	-	-	17.0	
LBP-TOP [11]	-	-	10.0	-
Motion mag+LBP [36]	0.2‡	**0.0**‡	14.4‡	-
Spectral cubes [12]	-	2.8	14.0	-
DMD [35]	5.3	3.8	21.8	-
Colour LBP [15]	0.4	2.9	6.2	-
Proposed method	**0.0**	3.5	**3.2**	**3.5**

‡ These extended results have been obtained from the technical report: https://repository.iiitd.edu.in/jspui/handle/123456789/138.

TABLE VI

COMPARISON BETWEEN THE PROPOSED COUNTERMEASURE AND STATE-OF-THE-ART METHODS ON THE THREE BENCHMARK DATASETS USING THE FRAME BASED EVALUATION

Method	Replay-Attack		CASIA	MSU
	EER	HTER	EER	EER
Motion [39]	11.6†	11.7†	26.6†	-
LBP [29]	13.9	13.8	18.2	
LBP-TOP [11]	7.9	7.6	-	-
CDD [10]	-	-	11.8	-
IQA[5]	-	15.2	32.4	-
CNN [56]	6.1	**2.1**	7.4	-
IDA [6]	-	7.4	-	8.5
Motion+LBP [46]	4.5	5.1	-	-
Proposed method	**0.4**	2.8	**2.1**	**4.9**

† These extended results have been obtained from [42].

combined. As we can see from Table IV, the combination of the features extracted from the HSV and YCbCr results in a significant performance enhancement, thus confirming the benefits of combining the different facial colour texture descriptions. The EER on the CASIA FASD and MSU MFSD has been reduced from 4.0% to 3.2% and from 4.9% to 3.5%, respectively, whereas the HTER on the Replay-Attack Database has been decreased from 4.3% to 3.3% (note that these peformance gains are computed by taking into account the best performances obtained using each colour space separately).

C. Comparison With the State of the Art

Tables V and VI provide a comparison with the state-of-the-art face spoofing detection techniques proposed in the literature. It can be seen from these tables that our proposed colour texture analysis based method outperforms the state-of-the-art algorithms on the two challenging CASIA FASD and MSU MFSD databases in both video and frame based evaluations. Our approach also achieves very competitive performance on the Replay-Attack Database. Most importantly, unlike most of the methods proposed in the literature, our proposed approach is able to achieve stable performance across all three benchmark datasets.

D. Cross-Database Analysis

In real-world applications, face anti-spoofing techniques are operating in open environments, thus unknown conditions and unseen attack scenarios will be faced. To gain insight into the generalization capabilities of our proposed method, we conducted a cross-database evaluation. To be more specific, the countermeasure was trained and tuned on one of the databases and then tested on the remaining two datasets, like performed in [42]. The obtained results are summarized in Table VII.

TABLE VII

THE PERFORMANCE OF THE CROSS-DATABASE EVALUATION IN TERMS OF HTER(%) ON THE CASIA FASD, MSU MFSD AND REPLAY-ATTACK

Test on:	CASIA		MSU		Replay-Attack		
Train on:	*Train*	*Test*	*Train*	*Test*	*Train*	*Dev*	*Test*
CASIA	-	-	20.0	20.8	28.8	32.3	30.0
MSU	43.3	48.7	-	-	30.5	35.8	35.6
Replay-Attack	36.3	39.2	33.8	34.5	-	-	-

When the countermeasure is trained on the CASIA FASD, we notice that the average of the HTER values on the different subsets is 20.4% for MSU MFSD and 30.3% for Replay-Attack Database. When the model is trained on Replay-Attack Database, the average HTER on CASIA FASD and MSU MFSD is 37.7% and 34.1%, respectively. When the cross-database performance is evaluated on CASIA FASD and Replay-Attack Database while training the model using the MSU MFSD, the average HTER is 46% and 33.9% for CASIA FASD and Replay-Attack Database, respectively. From these results, we observe that the models trained on MSU MFSD and Replay-Attack Database are not able to generalize as good as the model trained on the CASIA FASD. The reason behind this is that the CASIA FASD contains more variations in the collected data (e.g. imaging quality and proximity between the camera and the face) compared to the Replay-Attack and MSU databases. Therefore, the model optimized for these databases has difficulties to perform well in the new acquisition conditions. Even though the performance of our proposed approach decreases quite much compared to the intra-test, we see from Table VIII that the generalization capabilities are still much better compared to the state-of-the-art techniques.

Note that the results of the colour LBP descriptor shown in Table VIII are different from those previously reported in [15] in which the face bounding boxes were automatically extracted using KeyLemon's commercial software [57]. However, in this presented work, we were not able to use the same software with the MSU database. Thus, for consistency and fairness, we used the face locations provided by the authors of [6] who used PittPatt 5.2.2 SDK [54]) in their experiments with all the databases. The learnt lesson is that the face registration procedure affects the performance of the face anti-spoofing methods and especially in the cross-database scenario.

The investigations suggest that the facial colour texture representation seems be more stable in unknown conditions than texture descriptions extracted from gray-scale images. Thus, the use of colour texture information provides a way to improve the unsatisfying generalization capabilities of texture based approaches.

TABLE VIII
THE RESULTS OF THE CROSS-DATABASE EXPERIMENT ON THE CASIA FASD, REPLAY-ATTACK DATABASE AND MSU MFSD COMPARED WITH RELATED STUDIES

Method	Train:	Test:	Avg HTER
Motion [42]	CASIA	Replay	50.2
	Replay	CASIA	47.9
LBP [42]	CASIA	Replay	45.9
	Replay	CASIA	57.6
LBP-TOP [42]	CASIA	Replay	49.7
	Replay	CASIA	60.6
Motion-Mag [36]*	CASIA	Replay	50.1
	Replay	CASIA	47.0
Spectral cubes [12]	CASIA	Replay	34.4
	Replay	CASIA	50.0
CNN [56]	CASIA	Replay	48.5
	Replay	CASIA	45.5
LBP [15]**	CASIA	Replay	47.0
		MSU	36.6
	Replay	CASIA	39.6
		MSU	35.2
	MSU	CASIA	49.6
		Replay	42.0
Proposed method	CASIA	Replay	**30.3**
		MSU	20.4
	Replay	CASIA	**37.7**
		MSU	34.1
	MSU	CASIA	46.0
		Replay	33.9

* from https://repository.iiitd.edu.in/jspui/handle/123456789/138
** the results are obtained using the new face detector.

VI. CONCLUSIONS AND FUTURE WORK

In this article, we proposed to approach the problem of face anti-spoofing from the colour texture analysis point of view. We investigated how well different colour image representations (RGB, HSV and YCbCr) can be used for describing the intrinsic disparities in the colour texture between genuine faces and fake ones and if they provide complementary representations. The effectiveness of the different facial colour texture representations was studied by extracting different local descriptors from the individual image channels in the different colour spaces.

Extensive experiments on the three latest and most challenging spoofing databases (the CASIA FASD, the Replay-Attack Database and MSU MFSD) showed excellent results. On the CASIA FASD and the MSU MFSD, the proposed facial colour texture representation based on the combination CoLBP and LPQ features computed over HSV and YCbCr colour spaces outperformed the state of the art, while very competitive results were achieved on the Replay-Attack Database. More importantly, the proposed approach was able to achieve stable performance across all three benchmark datasets unlike most of the methods proposed in the literature. Furthermore, in our inter-database evaluation, the facial colour texture representation showed promising generalization capabilities, thus suggesting that colour texture seems to be more stable in unknown conditions compared to its gray-scale counterparts.

To be consistent with many previous studies, it is worth noting that face normalization or the limits of the face bounding box were not optimized in our experiments. However, these are shown to be important factors already in intra-database tests [6], [10], [56] and we noticed that they also significantly affect the cross-database performance. Improving the generalization capabilities of the colour texture analysis based face spoofing detection will be the main objective of our future work. Thus, we will study more closely how the size of the normalized face images and the used face bounding box and different face normalization methods affect both the intra-test and, especially, the inter-test performance. It is also of interest to investigate whether some feature descriptors or colour spaces lead to more robust and stable face representations across different acquisition conditions and spoofing scenarios. In addition, we aim to derive use case scenario specific facial colour representations and consider person-specific training for face spoofing detection.

REFERENCES

[1] Y. Li, K. Xu, Q. Yan, Y. Li, and R. H. Deng, "Understanding OSN-based facial disclosure against face authentication systems," in *Proc. 9th ACM Symp. Inf., Comput. Commun. Secur. (ASIA CCS)*, 2014, pp. 413–424.

[2] J. Li, Y. Wang, T. Tan, and A. K. Jain, "Live face detection based on the analysis of Fourier spectra," *Proc. SPIE*, vol. 5404, pp. 296–303, Aug. 2004.

[3] X. Tan, Y. Li, J. Liu, and L. Jiang, "Face liveness detection from a single image with sparse low rank bilinear discriminative model," in *Proc. 11th Eur. Conf. Comput. Vis., VI (ECCV)*, 2010, pp. 504–517.

[4] Z. Zhang, J. Yan, S. Liu, Z. Lei, D. Yi, and S. Z. Li, "A face antispoofing database with diverse attacks," in *Proc. 5th IAPR Int. Conf. Biometrics (ICB)*, Mar./Apr. 2012, pp. 26–31.

[5] J. Galbally and S. Marcel, "Face anti-spoofing based on general image quality assessment," in *Proc. IAPR/IEEE Int. Conf. Pattern Recognit. (ICPR)*, Aug. 2014, pp. 1173–1178.

[6] D. Wen, H. Han, and A. K. Jain, "Face spoof detection with image distortion analysis," *IEEE Trans. Inf. Forensics Security*, vol. 10, no. 4, pp. 746–761, Apr. 2015.

[7] J. Bai, T.-T. Ng, X. Gao, and Y.-Q. Shi, "Is physics-based liveness detection truly possible with a single image?" in *Proc. IEEE Int. Symp. Circuits Syst. (ISCAS)*, May/Jun. 2010, pp. 3425–3428.

[8] J. Määttä, A. Hadid, and M. Pietikäinen, "Face spoofing detection from single images using micro-texture analysis," in *Proc. Int. Joint Conf. Biometrics (IJCB)*, Oct. 2011, pp. 1–7.

[9] J. Maatta, A. Hadid, and M. Pietikäinen, "Face spoofing detection from single images using texture and local shape analysis," *IET Biometrics*, vol. 1, no. 1, pp. 3–10, Mar. 2012.

[10] J. Yang, Z. Lei, S. Liao, and S. Z. Li, "Face liveness detection with component dependent descriptor," in *Proc. IAPR Int. Conf. Biometrics (ICB)*, Jun. 2013, pp. 1–6.

[11] T. de Freitas Pereira *et al.*, "Face liveness detection using dynamic texture," *EURASIP J. Image Video Process.*, vol. 2014, no. 1, pp. 1–15, Dec. 2014.

[12] A. Pinto, H. Pedrini, W. R. Schwartz, and A. Rocha, "Face spoofing detection through visual codebooks of spectral temporal cubes," *IEEE Trans. Image Process.*, vol. 24, no. 12, pp. 4726–4740, Dec. 2015.

[13] *Digital Photography Review*, accessed on Nov. 1, 2015. [Online]. Available: http://www.dpreview.com/forums/thread/3861133

[14] N. Bonnier, "Contribution to spatial gamut mapping algorithms," M.S. thesis, Lab. Commun. Process. Inf. (LTCI), Télécom ParisTech, Paris, France, Sep. 2008.

[15] Z. Boulkenafet, J. Komulainen, and A. Hadid, "Face anti-spoofing based on color texture analysis," in *Proc. IEEE Int. Conf. Image Process. (ICIP)*, Sep. 2015, pp. 2636–2640.

[16] J. Y. Choi, K. Plataniotis, and Y. M. Ro, "Using colour local binary pattern features for face recognition," in *Proc. IEEE Int. Conf. Image Process. (ICIP)*, Sep. 2010, pp. 4541–4544.

[17] D. Gragnaniello, G. Poggi, C. Sansone, and L. Verdoliva, "An investigation of local descriptors for biometric spoofing detection," *IEEE Trans. Inf. Forensics Security*, vol. 10, no. 4, pp. 849–863, Apr. 2015.

[18] N. Erdogmus and S. Marcel, "Spoofing 2D face recognition systems with 3D masks," in *Proc. IEEE Int. Conf. Biometrics Special Interest Group*, Sep. 2013, pp. 1–8.

[19] I. Pavlidis and P. Symosek, "The imaging issue in an automatic face/disguise detection system," in *Proc. IEEE Workshop Comput. Vis. Beyond Visible Spectrum, Methods Appl. (CVBVS)*, 2000, pp. 15–24.

[20] Z. Zhang, D. Yi, Z. Lei, and S. Z. Li, "Face liveness detection by learning multispectral reflectance distributions," in *Proc. IEEE Int. Conf. Autom. Face Gesture Recognit. Workshops*, Mar. 2011, pp. 436–441.

[21] D. F. Smith, A. Wiliem, and B. C. Lovell, "Face recognition on consumer devices: Reflections on replay attacks," *IEEE Trans. Inf. Forensics Security*, vol. 10, no. 4, pp. 736–745, Apr. 2015.

[22] K. Kollreider, H. Fronthaler, M. I. Faraj, and J. Bigun, "Real-time face detection and motion analysis with application in 'liveness' assessment," *IEEE Trans. Inf. Forensics Security*, vol. 2, no. 3, pp. 548–558, Sep. 2007.

[23] E. S. Ng and A. Y. S. Chia, "Face verification using temporal affective cues," in *Proc. Int. Conf. Pattern Recognit. (ICPR)*, Nov. 2012, pp. 1249–1252.

[24] G. Chetty and M. Wagner, "Liveness verification in audio-video speaker authentication," in *Proc. 10th Austral. Int. Conf. Speech Sci. Technol.*, 2004, pp. 358–363.

[25] R. W. Frischholz and A. Werner, "Avoiding replay-attacks in a face recognition system using head-pose estimation," in *Proc. IEEE Int. Workshop Anal. Modeling Faces Gestures*, Oct. 2003, pp. 234–235.

[26] M. De Marsico, M. Nappi, D. Riccio, and J.-L. Dugelay, "Moving face spoofing detection via 3D projective invariants," in *Proc. 5th IAPR Int. Conf. Biometrics (ICB)*, Mar./Apr. 2012, pp. 73–78.

[27] T. Wang, J. Yang, Z. Lei, S. Liao, and S. Z. Li, "Face liveness detection using 3D structure recovered from a single camera," in *Proc. IAPR Int. Conf. Biometrics (ICB)*, Jun. 2013, pp. 1–6.

[28] G. Pan, L. Sun, Z. Wu, and Y. Wang, "Monocular camera-based face liveness detection by combining eyeblink and scene context," *Telecommun. Syst.*, vol. 47, nos. 3–4, pp. 215–225, 2011.

[29] I. Chingovska, A. Anjos, and S. Marcel, "On the effectiveness of local binary patterns in face anti-spoofing," in *Proc. Int. Conf. Biometrics Special Interest Group (BIOSIG)*, Sep. 2012, pp. 1–7.

[30] N. Erdogmus and S. Marcel, "Spoofing in 2D face recognition with 3D masks and anti-spoofing with Kinect," in *Proc. IEEE 6th Int. Conf. Biometrics, Theory, Appl. Syst. Conf. (BTAS)*, Sep. 2013, pp. 1–6.

[31] M. M. Chakka *et al.*, "Competition on counter measures to 2-D facial spoofing attacks," in *Proc. IAPR, IEEE Int. Joint Conf. Biometrics (IJCB)*, Oct. 2011, pp. 1–6.

[32] I. Chingovska *et al.*, "The 2nd competition on counter measures to 2D face spoofing attacks," in *Proc. IAPR, Int. Conf. Biometrics (ICB)*, Jun. 2013.

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

[34] G. Pan, Z. Wu, and L. Sun, "Liveness detection for face recognition," in *Recent Advances in Face Recognition*. Vukovar, Croatia: InTech, 2008, pp. 109–124.

[35] S. Tirunagari, N. Poh, D. Windridge, A. Iorliam, N. Suki, and A. T. S. Ho, "Detection of face spoofing using visual dynamics," *IEEE Trans. Inf. Forensics Security*, vol. 10, no. 4, pp. 762–777, Apr. 2015.

[36] S. Bharadwaj, T. I. Dhamecha, M. Vatsa, and R. Singh, "Computationally efficient face spoofing detection with motion magnification," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. Workshop*, Jun. 2013, pp. 105–110.

[37] W. Bao, H. Li, N. Li, and W. Jiang, "A liveness detection method for face recognition based on optical flow field," in *Proc. Int. Conf. Image Anal. Signal Process.*, Apr. 2009, pp. 233–236.

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

[39] A. Anjos and S. Marcel, "Counter-measures to photo attacks in face recognition: A public database and a baseline," in *Proc. IAPR IEEE Int. Joint Conf. Biometrics (IJCB)*, Oct. 2011, pp. 1–7.

[40] J. Yan, Z. Zhang, Z. Lei, D. Yi, and S. Z. Li, "Face liveness detection by exploring multiple scenic clues," in *Proc. 12th Int. Conf. Control Autom. Robot. Vis. (ICARCV)*, Dec. 2012, pp. 188–193.

[41] N. Kose and J.-L. Dugelay, "Countermeasure for the protection of face recognition systems against mask attacks," in *Proc. 10th Int. Conf. Autom. Face Gesture Recognit.*, Shanghai, China, Apr. 2013, pp. 1–6.

[42] T. de Freitas Pereira, A. Anjos, J. M. De Martino, and S. Marcel, "Can face anti-spoofing countermeasures work in a real world scenario?" in *Proc. Int. Conf. Biometrics (ICB)*, Jun. 2013, pp. 1–8.

[43] I. Chingovska and A. R. dos Anjos, "On the use of client identity information for face antispoofing," *IEEE Trans. Inf. Forensics Security*, vol. 10, no. 4, pp. 787–796, Apr. 2015.

[44] J. Yang, Z. Lei, D. Yi, and S. Z. Li, "Person-specific face antispoofing with subject domain adaptation," *IEEE Trans. Inf. Forensics Security*, vol. 10, no. 4, pp. 797–809, Apr. 2015.

[45] J. Komulainen, A. Hadid, and M. Pietikäinen, "Context based face anti-spoofing," in *Proc. IEEE 6th Int. Conf. Biometrics, Theory, Appl. Syst. (BTAS)*, Sep./Oct. 2013, pp. 1–8.

[46] J. Komulainen, A. Hadid, M. Pietikäinen, A. Anjos, and S. Marcel, "Complementary countermeasures for detecting scenic face spoofing attacks," in *Proc. IAPR Int. Conf. Biometrics*, Jun. 2013, pp. 1–7.

[47] I. Chingovska, A. Anjos, and S. Marcel, "Anti-spoofing in action: Joint operation with a verification system," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. Workshop Biometrics*, Jun. 2013, pp. 98–104.

[48] T. Ahonen, A. Hadid, and M. Pietikäinen, "Face description with local binary patterns: Application to face recognition," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 28, no. 12, pp. 2037–2041, Dec. 2006.

[49] R. Lukac and K. N. Plataniotis, *Color Image Processing: Methods and Applications*, vol. 8. New York, NY, USA: CRC Press, 2007.

[50] T. Ojala, M. Pietikäinen, and T. Mäenpaa, "Multiresolution gray-scale and rotation invariant texture classification with local binary patterns," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 24, no. 7, pp. 971–987, Jul. 2002.

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

[52] V. Ojansivu and J. Heikkilä, "Blur insensitive texture classification using local phase quantization," in *Image and Signal Processing* (Lecture Notes in Computer Science), vol. 5099. Berlin, Germany: Springer, 2008, pp. 236–243.

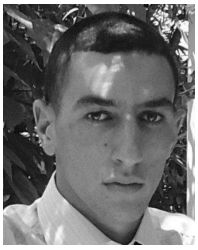
[53] J. Kannala and E. Rahtu, "BSIF: Binarized statistical image features," in *Proc. 21st Int. Conf. Pattern Recognit. (ICPR)*, Nov. 2012, pp. 1363–1366.

[54] Pittsburgh Pattern Recognition (PittPatt), accessed on Nov. 1, 2015. *PittPatt Software Developer Kit acquired by Google*. [Online]. Available: http://www.pittpatt.com/

[55] 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.*, vol. 9, pp. 1871–1874, Jun. 2008.

[56] J. Yang, Z. Lei, and S. Z. Li. (2014). "Learn convolutional neural network for face anti-spoofing." [Online]. Available: http://arxiv.org/abs/1408.5601

[57] *KeyKemon*, accessed on Nov. 1, 2015. [Online]. Available: http://www.keylemon.com/



Zinelabidine Boulkenafet received the Engineering degree and the master's degree in computer science from the National School of Computer Science, Algiers, Algeria. He is currently pursuing the Ph.D. degree with the Center for Machine Vision and Signal Analysis, University of Oulu, Finland. His research interests include signal and image processing, biometrics, and spoofing detection.

Jukka Komulainen received the M.Sc. and D.Sc. degrees in information engineering from the University of Oulu, in 2010 and 2015, respectively. He holds a postdoctoral position with the Center for Machine Vision and Signal Analysis, University of Oulu. His research interests include image processing, machine learning, and pattern recognition, with a particular focus on biometric authentication and especially anti-spoofing.

Abdenour Hadid received the D.Sc. degree in electrical and information engineering from the University of Oulu, Finland, in 2005. He is currently an Academy Research Fellow with the Center for Machine Vision and Signal Analysis, University of Oulu. He is regularly visiting the School of Electronics and Information, Northwestern Polytechnical University, Xi'an, China. His research interests include computer vision, machine learning, and pattern recognition with a particular focus on biometrics.