Face Spoof Detection With Image Distortion Analysis

Di Wen, *Member, IEEE*, Hu Han, *Member, IEEE*, and Anil K. Jain, *Life Fellow, IEEE*

***Abstract*— Automatic face recognition is now widely used in applications ranging from deduplication of identity to authentication of mobile payment. This popularity of face recognition has raised concerns about face spoof attacks (also known as biometric sensor presentation attacks), where a photo or video of an authorized person's face could be used to gain access to facilities or services. While a number of face spoof detection techniques have been proposed, their generalization ability has not been adequately addressed. We propose an efficient and rather robust face spoof detection algorithm based on image distortion analysis (IDA). Four different features (specular reflection, blurriness, chromatic moment, and color diversity) are extracted to form the IDA feature vector. An ensemble classifier, consisting of multiple SVM classifiers trained for different face spoof attacks (e.g., printed photo and replayed video), is used to distinguish between genuine (live) and spoof faces. The proposed approach is extended to multiframe face spoof detection in videos using a voting-based scheme. We also collect a face spoof database, MSU mobile face spoofing database (MSU MFSD), using two mobile devices (Google Nexus 5 and MacBook Air) with three types of spoof attacks (printed photo, replayed video with iPhone 5S, and replayed video with iPad Air). Experimental results on two public-domain face spoof databases (Idiap REPLAY-ATTACK and CASIA FASD), and the MSU MFSD database show that the proposed approach outperforms the state-of-the-art methods in spoof detection. Our results also highlight the difficulty in separating genuine and spoof faces, especially in cross-database and cross-device scenarios.**

***Index Terms*— Face recognition, spoof detection, image distortion analysis, ensemble classifier, cross-database, cross-device.**

I. Introduction

As a convenient user authentication technique, automatic face recognition has attracted increasing attention in various access control applications, especially for mobile phone unlocking. With the release of face unlocking functionality in the Android mobile operating system, face recognition becomes another biometric authentication technique for mobile phones, similar to fingerprint authentication (Touch ID) in the iOS system. Unlike fingerprint authentication, face recognition does not require any additional sensor since all smart phones come equipped with a front facing camera. However, similar to other biometric modalities [1], [2], we need to address concerns about face spoof attacks on face recognition systems, particularly in unconstrained sensing and uncooperative subject scenarios [3].

It is relatively easier to acquire a person's face image or video (e.g., with a digital camera or from social media) than it is to acquire other biometric traits such as fingerprint, palm print, and iris. Further, the cost of launching a face spoof attack, such as a printed photo, displayed photo, or replayed video is relatively low (see Fig. 1). State of the art Commercial Off-The-Shelf (COTS) face recognition systems are not well designed to differentiate spoof faces from genuine live faces. Figure 2 shows the face identification performance of

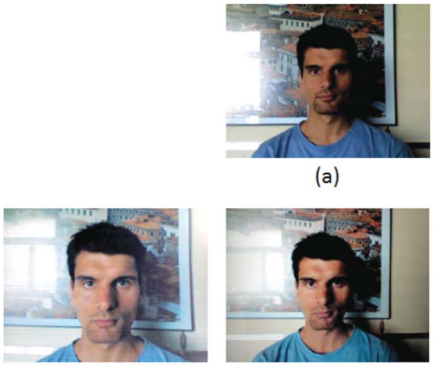


Fig. 1. A genuine face image (a) of a subject in the Idiap databases [4], [5] and three examples of spoofs of the same subject using a (b) printed photo, (c) displayed photo (on a tablet screen), and (d) 3D face mask.

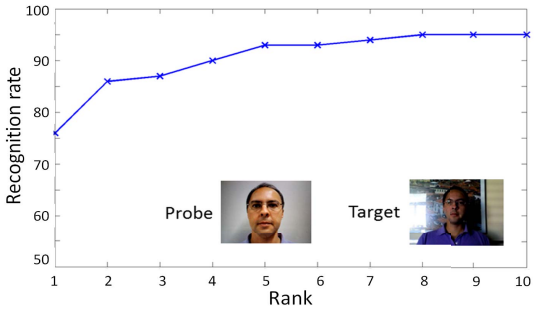


Fig. 2. COTS face recognition systems can not well distinguish between genuine and spoof faces. Recognition rate of a state-of-the-art COTS face recognition system in matching spoof probe face images against genuine gallery face images. Images in this experiment come from the Idiap REPLAY-ATTACK database [4]. The gallery (target) set contains 50 videos from 50 subjects in the database, and the probe set contains 100 replayed videos from the same 50 subjects captured at a different location than the gallery videos.

Manuscript received June 30, 2014; revised October 27, 2014 and January 19, 2015; accepted January 22, 2015. Date of publication February 4, 2015; date of current version March 13, 2015. This work was supported by the Center for Identification Technology Research under Grant 14S-04W-12. The associate editor coordinating the review of this manuscript and approving it for publication was Prof. Stan Z. Li. *(Corresponding author: Hu Han.)*

D. Wen was with the Department of Computer Science and Engineering, Michigan State University, East Lansing, MI 48824 USA. He is now with the Department of Biomedical Engineering, Case Western Reserve University, Cleveland, OH 44106 USA (e-mail: wendi@msu.edu).

H. Han and A. K. Jain are with the Department of Computer Science and Engineering, Michigan State University, East Lansing, MI 48824 USA (e-mail: hhan@msu.edu; jain@msu.edu).

This paper has supplementary downloadable material at http://ieeexplore.ieee.org, provided by the authors. The file consists of a supplementary video. The material is 11 MB in size.

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.2015.2400395

a COTS face recognition system (COTS1[1]) when spoof faces as probe are matched to genuine faces in the gallery. In this experiment, more than 70% of probe videos (spoof faces) were *successfully* matched to the gallery mates by COTS1 at rank-1, indicating that COTS1 cannot effectively distinguish between genuine and spoof faces. In this paper we do not address 3D face mask attacks, which are more expensive to launch.[2] Instead, we focus on printed photo and replayed video attacks.

The fragility of face recognition systems to face spoof attacks has motivated a number of studies on face spoof detection [4], [7]–[12]. However, published studies are limited in their scope because the training and testing images (videos) used were captured under the same imaging conditions. It is essential to develop robust and efficient face spoof detection (or anti-spoofing) algorithms that generalize well to new imaging conditions and environments. In this paper, we study the cross-database face spoof detection problem and propose a face spoof detection approach based on *Image Distortion Analysis* (IDA). The contributions of this paper can be summarized as follows:

i) A face spoof detection algorithm based on IDA, which is effective in grasping the intrinsic distortions of spoof face images with respect to the genuine face images.
ii) We construct a face spoof database, named the MSU Mobile Face Spoof Database (MSU MFSD), using the cameras of a laptop (MacBook Air[3]) and a mobile phone (Google Nexus 5[4]) and three types of attack medium (iPad, iPhone, and printed photo). The MSU MFSD database allows us to evaluate the generalization ability of face spoof detection algorithms across different cameras and illumination conditions with mobile devices. For a subset of the MSU MFSD database (35 subjects), we have the subjects' permission to make their data publicly available.
iii) We present results for both intra-database and cross-database scenarios using two public-domain face spoof databases (Idiap REPLAY-ATTACK and CASIA FASD), and our own database (MSU MFSD).

The remainder of the paper is organized as follows. Prior work on face spoof detection is reviewed in Section II. The proposed feature types for IDA are detailed in Section III. The genuine vs. spoof face discrimination method is described in Section IV. Testing protocols and baseline methods are given in Section V. In Section VI, we summarize the public-domain face spoof databases and describe the MSU MFSD database that we have collected. Experimental results are presented in Section VII. Finally, we conclude this work and discuss future work in Section VIII.

II. Prior Work

To our knowledge, one of the earliest studies on face spoof detection was reported in 2004 by Li et al. [13]. With the growing popularity of using face recognition for access control, this topic has attracted significant attention over the past five years [4], [7]–[12]. One of the major focus of the FP7 EU funded project, TABULA RASA [14], is "trusted biometrics under spoofing attacks". Here, we provide a brief summary of face spoof detection algorithms published in the literature along with their strengths and limitations in terms of (i) robustness and generalization ability, and (ii) real-time response and usability. According to different types of cues used in face spoof detection, published methods can be categorized into four groups: (i) *motion based methods*, (iii) *texture based methods*, (iii) *method based on image quality analysis*, and (iv) *methods based on other cues*.

(i) *Motion Based Methods:* These methods, designed primarily to counter printed photo attacks, capture a very important cue for vitality: the subconscious motion of organs and muscles in a live face, such as eye blink [10], mouth movement [15] and head rotation [11]. Given that motion is a relative feature across video frames, these methods are expected to have better generalization ability than the texture based methods that will be discussed below. However, the limitations of motion based methods are apparent. The frequency of facial motion is restricted by the human physiological rhythm, which ranges from 0.2 to 0.5 Hz [12]. Therefore, it takes a relatively long time (usually > 3s) to accumulate stable vitality features for face spoof detection. Additionally, motion based methods can be easily circumvented or confused by other motions, e.g., background motion, that are irrelevant to facial liveness or replayed motion in the video attacks.

(ii) *Texture Based Methods:* To counter both the printed photo and replayed video attacks, texture based methods were proposed to extract image artifacts in spoof face images. In [18], the authors argued that texture features (like LBP, DoG, or HOG) are capable of differentiating artifacts in spoof faces from the genuine faces. Texture based methods have achieved significant success on the Idiap and CASIA databases. The Half Total Error Rate (HTER)[5] on the Idiap database was reduced from 13.87% in [4] and 7.60% in [16] to 6.62% in [12] by incorporating texture cues. Unlike motion based methods, texture based methods need only a single image to detect a spoof. However, the generalization ability of many texture based methods has been found to be poor. A study reported in [17] showed that for two of the texture based methods (proposed in [4] and [16]), the HTER increased dramatically under the cross-database scenarios (where the training and testing sets came from different face spoof databases). Due to the intrinsic data-driven nature of texture based methods, they can be easily over-fitted to one particular illumination and imagery condition and hence do not generalize well to databases collected under different conditions.

[1]We denote the COTS system used in this paper as COTS1 in order to make it anonymous.

[2]A public domain database of 3D face mask attacks is available [5] and a few publications address measures to detect such attacks [5], [6].

[3]Model specification: http://support.apple.com/kb/SP670

[4]http://www.google.com/nexus/5/

[5]HTER is the average of False Rejection Rate and False Acceptance Rate at the threshold determined in the training set [7].

TABLE I
A COMPARISON OF DIFFERENT FACE SPOOF DETECTION METHODS

Method	Strengths	Limitations	State-of-the-art performance
Motion based methods [10]–[12], [15]	• Good generalization ability	• Low robustness (can be circumvented by fake motion) • Slow response ($>$ 3s) • High computational complexity (image registration needed)	*Intra-DB* [12]: HTER = 1.25% for Idiap REPLAY-ATTACK
Texture based methods [4], [8], [12], [16]–[18]	• Fast response ($<$ 1s) • Low computational complexity	• Poor generalization ability (vulnerable to the variations in acquisition conditions)	*Intra-DB* [16]: HTER = 7.60% for Idiap REPLAY-ATTACK *Intra-DB* [18]: EER = 11.8% for CASIA FASD *Intra-DB* [12]: HTER = 6.62% for Idiap REPLAY-ATTACK
Methods based on other cues [6], [19]–[21]	• High robustness	• Additional sensing or processing technique needed (IR, audio, 3D, etc.) • Slow response ($>$ 3s) when using audio and 3D cues	Intra-DB [21]: EER = 8.06% for VidTIMIT EER = 9.23% for DaFEx
Image quality analysis based methods [22], and the proposed method	• Good generalization ability • Fast response ($<$ 1s) • Low computational complexity	• Different classifiers needed for different spoof attacks	*Intra-DB*: TPR = 92.2% @ FAR = 10% for Idiap REPLAY-ATTACK *Cross-DB*: TPR = 75.5% @ FAR = 10% for MSU MFSD

(iii) *Image Quality Analysis Based Methods:* A recent work [22] proposed a biometric liveness detection method for iris, fingerprint and face images using 25 image quality measures, including 21 full-reference measures and 4 non-reference measures. Compared to [22], our work is different in the following aspects: (1) While 25 features are required in [22] to get good results, no face-specific information has been considered in designing informative features for face spoof detection. On the contrary, four features are designed specifically for face feature representation in our method, and we demonstrate the effectiveness of these features for spoof face detection. (2) While the authors of [22] evaluated their method on *only* the Idiap-Replay database, we have used both the Idiap and CASIA databases, which are two important public-domain databases. (3) While the work in [22] aims at designing a generic liveness detection method across different biometric modalities, the training and testing of each modality were still performed under intra-database scenarios (same database for training and testing, even though cross-validation is used). By contrast, the proposed approach aims to improve the generalization ability under cross-database scenarios, which has seldom been explored in the biometrics community.

(iv) *Methods Based on Other Cues:* Face spoof countermeasures using cues derived from sources other than 2D intensity image, such as 3D depth [19], IR image [6], spoofing context [20], and voice [21] have also been proposed. However, these methods impose extra requirements on the user or the face recognition system, and hence have a narrower application range. For example, an IR sensor was required in [6], a microphone and speech analyzer were required in [21], and multiple face images taken from different viewpoints were required in [19]. Additionally, the spoofing context method proposed in [20] can be circumvented by concealing the spoofing medium.

Table I compares these four types of spoof detection methods. These four types of methods can also be combined to utilize multiple cues for face spoof detection. For example, in [12], the authors showed that appropriately magnified motion cue [23] improves the performance of texture based approaches (HTER $=$ 6.62% on the Idiap database with motion magnification compared to HTER $=$ 11.75% without motion magnification, both using LBP features). The authors also showed that combining the Histogram of Oriented Optical Flow (HOOF) feature with motion magnification achieved the best performance on the Idiap database (HTER $=$ 1.25%). However, motion magnification, limited by human physiological rhythm, cannot reach the reported performance [12] without accumulating a large number of video frames ($>$200 frames), making these methods unsuitable for real-time response.

Though a number of face spoof detection methods have been reported, to our knowledge, none of them generalizes well to cross-database scenarios [17]. In particular, there is a lack of investigation on how face spoof detection methods perform in cross-database scenarios. The fundamental differences between intra-database and cross-database scenarios are as follows:

i) In an intra-database scenario, it is assumed that the spoof media (e.g., photo and screen display), camera, environmental factors, and even the subjects are known to a face liveness detection system. This assumption does not hold in most of the real scenarios. The intra-database performance of a face liveness detection system is only the upper bound in terms of performance that cannot be expected in real applications.

ii) In cross-database scenario, we permit differences of spoof media, cameras, environments, and subjects during the system development stage and the system deployment stage. Hence this cross-database performance better reflects the actual performance of a system that can be expected in real applications.

iii) Existing methods, particularly methods using texture features, commonly used features (e.g., LBP) that are capable of capturing facial details and differentiating one subject from the other (for the purpose of face recognition). As a result, when the same features are used to differentiate a genuine face from a spoof face, they either contain some redundant information for liveness detection or information that is too person specific. These two factors limit the generalization ability of existing methods. To solve this problem, we have proposed a feature set based on Image Distortion Analysis (IDA) with real-time response (extracted from a single image with efficient computation) and better generalization performance in the cross-database scenario. Compared to the existing methods, the proposed method does not try to extract features that capture the facial details, but try to capture the face image quality differences due to the different reflection properties of different materials, e.g., facial skin, paper, and screen. As a result, experimental results show that the proposed method has better generalization ability.

III. Features Derived From Image Distortion Analysis

In mobile applications, the real-time response of face spoof detection requires that a decision be made based on a limited number of frames, e.g., no more than 30 frames ($\sim$1 sec. for videos of 30 fps). Therefore, we aim to design discriminative features that are capable of differentiating between genuine and spoof faces based on a single frame.

Given a scenario where a genuine face or a spoof face (such as a printed photo or replayed video on a screen) is presented to a camera in the same imaging environment, the main difference between genuine and spoof face images is due to the "shape" and characteristics of the facial surface in front of the camera. According to the Dichromatic Reflection Model [24], light reflectance I of an object at a specific location x can be decomposed into the following diffuse reflection (I_d) and specular reflection (I_s) components:

$$I(x) = I_d + I_s = w_d(x)S(x)E(x) + w_s(x)E(x) \qquad (1)$$

where $E(x)$ is the incident light intensity, $w_d(x)$ and $w_s(x)$ are the geometric factors for the diffuse and specular reflections, respectively, and $S(x)$ is the local diffuse reflectance ratio. Since 2D spoof faces are recaptured from original genuine face images, the formation of spoof face image intensity $I'(x)$ can be modeled as follows:

$$I'(x) = I'_d + I'_s = F(I(x)) + w'_s(x)E'(x) \qquad (2)$$

Note that equation (1) and (2) only model the reflectance difference between genuine and spoof faces and have not considered the final image quality after camera capture. In equation (2), we substitute the diffuse reflection of spoof face image I'_d by $F(I(x))$ because the diffuse reflection is determined by the distorted transformation of the original face image $I(x)$. Therefore, the total distortion in $I'(x)$ compared to $I(x)$ consists of two parts: i) distortion in the diffuse reflection component (I'_d), and ii) distortion in the specular reflection component (I'_s), both of which are related to the spoofing medium. In particular, I'_d is correlated with the original face image $I(x)$, while I'_s is independent of $I(x)$. Furthermore, the distortion function $F(\cdot)$ in the diffuse reflectance component can be modeled as

$$F(I(x)) = H(G \circledast I(x)), \qquad (3)$$

where $G(\cdot)$ is a low pass point spread function (blurring the original face image) and $H(\cdot)$ is a histogram transformation function (distorting color intensity). Explanation of $G(\cdot)$ and $H(\cdot)$ in printed photo attack and replay video attack is detailed below. Based on this imaging model, we provide an analysis of the significant differences between genuine faces and two types of spoof faces (printed photo and replay video or photo attacks) studied in this paper.

1) *Printed Photo Attack:* In printed photo attack, $I(x)$ is first transformed to the printed ink intensity on the paper and then to the final image intensity through diffusion reflection from the paper surface. During this transformation, $G(\cdot)$ and $H(\cdot)$ are determined by the printer frequency and chromatic fidelity. For high resolution color printer, the distortion of $G(\cdot)$ can be neglected, but not for $H(\cdot)$, since it has been reported that the color printing process usually reduces the image contrast [25]. Therefore, image distortion in printed photo attack can be approximated by a contrast degrading transformation.
2) *Replay Video Attack:* In replay video attack, $I(x)$ is transformed to the radiating intensity of pixels on LCD screen. Therefore, $G(\cdot)$ is determined by the frequency band width of the LCD panel, the distortion of which can be neglected. $H(\cdot)$ is related to the LCD color distortion and intensity transformation properties.

Besides the difference in diffuse reflectance, the specular reflectance of the spoof face also differentiates from that of the genuine face, which is caused by the spoof medium surface. Due to the glassy surface of tablet/mobile phone and the glossy ink layer on the printed paper, there is usually a specular reflection around the spoof face image. While for a genuine 3D face, specular reflection is only located in specific fiducial locations (such as nose tip, glasses, forehead, cheeks, etc.). Therefore, pooling the specular reflection from the entire face image can also capture the image distortion in spoof faces.

Besides the above distortions in the reflecting process, there is also distortion introduced by the capturing process. Although the capturing distortion can apply to both genuine and spoof faces. The spoof faces are more vulnerable to such distortion because they are usually captured in close distance to conceal the discontinuity of spoof medium frame. For example, defocused blurriness is commonly seen in both printed photo and replayed video attacks.

Based on the above analysis, the major distortions in a spoof face image include: (1) specular reflection from the printed paper surface or LCD screen; (2) image blurriness due to camera defocus; (3) image chromaticity and contrast distortion due to imperfect color rendering of printer or LCD screen; and (4) color diversity distortion due to limited color resolution of printer or LCD screen.

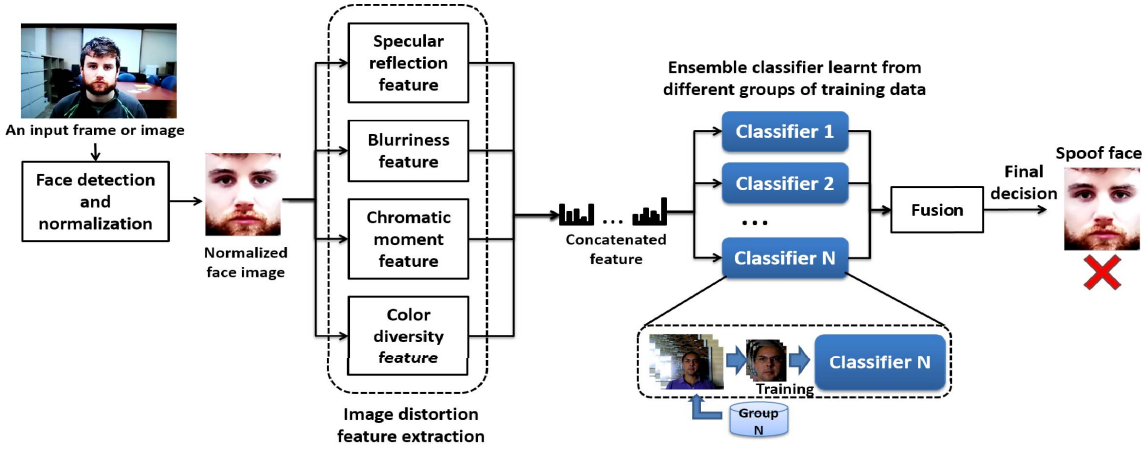


Fig. 3. The proposed face spoof detection algorithm based on Image Distortion Analysis.

There might be some other distortions present in spoof face images such as geometric distortion (e.g., paper warping) and artificial texture patterns (e.g., Moiré pattern); however, these distortions are camera and illumination dependent. For example, geometric distortion varies with illumination and the artificial texture pattern can only be discerned by a high quality camera. Hence, we focus only on the above four general sources of image distortion in spoof face images (specular reflection, blurriness, chromaticity, and color diversity), and design the corresponding features for face spoof detection.

Figure 3 shows the system diagram of the proposed spoof detection algorithm based on IDA. The input face image is first aligned based on two eyes locations and normalized to 144 × 120 pixels with an interpupillary distance (IPD) of 60 pixels. For face detection and eye localization, we use the PittPatt 5.2.2 SDK [26], which works successfully for about 99% of the faces in the Idiap, CASIA, and MSU face spoof databases. Our experiments show that face alignment and cropped face size are very important for spoof detection because they significantly reduces the influences of facial and background variations that are irrelevant to spoof detection. For each normalized face image, four different IDA features are extracted, constituting a 121-dimensional feature vector. This feature vector is then fed into multiple SVM classifiers, each trained on a different group of spoof training samples (e.g., printed photo attack and replayed video attack). The classifier outputs are fused to give the final binary decision (ensemble classification): genuine or spoof face.

A. Specular Reflection Features

Specular reflection component image has been widely used for specular reflection removal [27] and face illumination normalization [28]. In this paper, we separate the specular reflection component I_s from an input face image or video frame utilizing an iterative method (with 6 iterations) proposed in [29], which assumes that the illumination is i) from a single source, ii) of uniform color, and iii) not over-saturated. Given that most of the face images (in the Idiap, CASIA, and MSU databases) are captured indoors under relatively controlled illumination, these three assumptions are reasonable.[6] Figures 4 (a,b) illustrate the difference between the specular reflection components extracted from a genuine face and the corresponding spoof face.

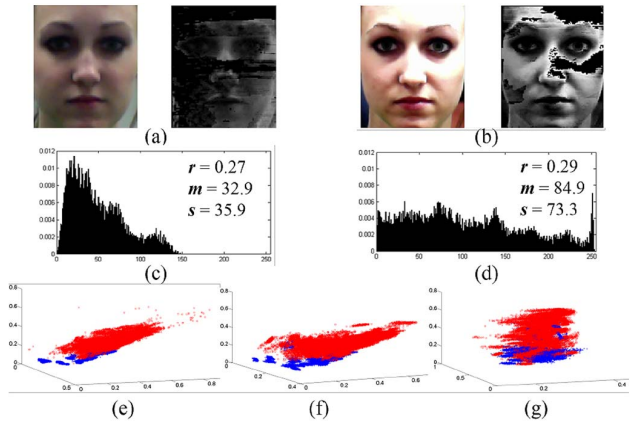


Fig. 4. Illustration of specular reflection features. (a) A genuine face image and the detected specular reflection component; (b) A spoof face (replayed video) and the detected specular reflection component; (c-d) histograms and specific feature values of the specular reflection components in (a) and (b), respectively; (e-g) distributions of the three specular reflection features (blue: genuine samples, red: spoof samples) in the Idiap training, Idiap testing, and MSU testing sets, respectively.

After calculating the specular reflection component image I_s, we represent the specularity intensity distribution with three dimensional features: i) specular pixel percentage r, ii) mean intensity of specular pixels μ, and iii) variance of specular pixel intensities σ.

However, as argued in [32], the method in [29] extracts specular components based on chromatic difference analysis, which often incorrectly classifies the mono-chromatic regions as specular components. To correct such errors, we exclude the high-intensity mono-chromatic pixels in I_s from specular components (as in [32]). Specifically, only pixels in the intensity range $(1.5\mu, 4\mu)$ are counted as specular pixels.

[6]To extract reliable specular reflection features under uncontrolled conditions, one possible method is to utilize more robust illumination estimation models, such as the multi-clue based illumination estimation method [30] or the illumination-reflectance model [31].

Figures 4 (a-d) show the three dimensional specular reflection features calculated for a genuine and a spoof face of a subject in the MSU database. Figures 4 (e-g) visualize the 3D distributions of the specular reflection features of genuine and spoof faces in the Idiap training, Idiap testing and MSU testing datasets. These distributions suggest that using the specular reflection feature, a classifier trained on the Idiap training set can achieve good performance on both the Idiap and MSU testing sets.

B. Blurriness Features

For short distance spoof attacks, spoof faces are often defocused in mobile phone cameras. The reason is that the spoofing medium (printed paper, tablet screen, and mobile phone screen) usually have limited size, and the attackers have to place them close to the camera in order to conceal the boundaries of the attack medium. As a result, spoof faces tend to be defocused, and the image blur due to defocus can be used as another cue for anti-spoofing.

We utilize two types of blurriness features (denoted as b_1 and b_2) that were proposed in [33] and [34], respectively. In [33], blurriness is measured based on the difference between the original input image and its blurred version. The larger the difference, the lower the blurriness in the original image. In [34], blurriness is measured based on the average edge width in the input image. Both these two methods output non-reference (without a clear image as reference) blurriness score between $0 \sim 1$, but emphasizing different measures of blurriness.

C. Chromatic Moment Features

Recaptured face images tend to show a different color distribution compared to colors in the genuine face images. This is caused by the imperfect color reproduction property of printing and display media. This chromatic degradation was explored in [35] for detecting recaptured images, but its effectiveness in spoof face detection is unknown. Since the absolute color distribution is dependent on illumination and camera variations, we propose to devise invariant features to detect abnormal chromaticity in spoof faces. That is, we first convert the normalized facial image from the RGB space into the HSV (Hue, Saturation, and Value) space and then compute the mean, deviation, and skewness of each channel as a chromatic feature. Since these three features are equivalent to the three statistical moments in each channel, they are also referred to as chromatic moment features. Besides these three features, the percentages of pixels in the minimal and maximal histogram bins of each channel are used as two additional features. So the dimensionality of the chromatic moment feature vector is $5 \times 3 = 15$. Figure 5 illustrates the presence of color distortion in a spoof face.

D. Color Diversity Features

Another important difference between genuine and spoof faces is the color diversity. In particular, genuine faces tend to have richer colors. This diversity tends to fade out in spoof

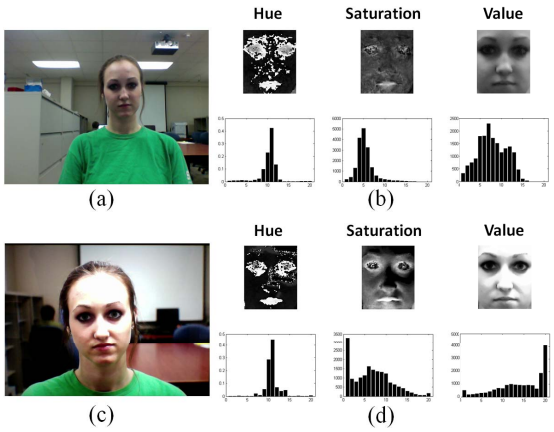


Fig. 5. Examples of chromatic difference between a genuine face and a spoof face. (a) and (c): The genuine face and spoof face images; (b) and (d): Hue, Saturation, and Value component images (top row) and their histograms (bottom row). The abnormality of the histogram for the spoof face can be measured by the three chromatic moments.

faces due to the color reproduction loss during image/video recapture. In this paper, we follow the method used in [35] to measure the image color diversity. First, color quantization (with 32 steps in the red, green and blue channels, respectively) is performed on the normalized face image. Two measurements are then pooled from the color distribution: i) the histogram bin counts of the top 100 most frequently appearing colors, and ii) the number of distinct colors appearing in the normalized face image. The dimensionality of the color diversity feature vector is 101.

The above four types of feature (specular reflection, blurriness, chromatic moment, and color diversity) are finally concatenated together, resulting in an IDA feature vector with 121 dimensions. Although the IDA feature vector is extracted from the facial region, it contains only image distortion information, and not any characterization of facial appearance. Therefore, we expect that the IDA feature can alleviate the problem of training bias encountered in the commonly used texture features.

IV. Classification Method

A. Ensemble Classifier

Given that our aim is to design an efficient face spoof detection system with good generalization ability and quick response, it is desirable to have an efficient classifier for the extracted IDA features. Following the success of SVM [36] in signal processing [37], pattern recognition and classification applications [38], [39], we choose to use SVM via the LibSVM Library [40]. There are also a number of variations of SVM for handling large-scale classification problems, such as LIBLINEAR [41] and ALM-SVM [42]; however, most of the public-domain face spoof databases (including the databases used in our experiments) are of limited size in terms of the number of still images, video tracks, and subjects. A SVM classifier with RBF kernel is trained for each group of training data, with parameters optimized by cross-validation.

On the other hand, it is understood that different spoof attacks will have different sample distributions in the

IDA feature space. For example, while the printed attack samples tend to have lower contrast than the genuine samples, the replay attack samples tend to have higher contrast. Different types of attacks might also have different chromatic distortion characteristics. Therefore, instead of training a single binary classifier, an ensemble classifier is more appropriate to cover various spoof attacks.

In this paper, we propose to use an ensemble classifier scheme by training multiple constituent spoof classifiers in different groups of spoof attack samples. For a specific spoof database, we construct separate groups of training samples as follows: First, the spoof samples are divided into K groups according to the attack type. Second, a specific training set is constructed by combining all genuine samples and a single group of spoof samples, resulting in K training sets. In our experiments, we find that by training two constituent classifiers ($K = 2$) on two groups of spoof attacks separately, i.e., printed attack and replay attack, the ensemble classifier performs better than training a single classifier on the whole database.

In the testing stage, the input feature vector is fed to all constituent classifiers and their outputs are fused to get the final result. We have evaluated two types of score-level fusion schemes: *sum* and *min* rules [43]. In most cases, the *min* rule performs better and so it is used in all our experiments below.

B. Multi-Frame Fusion

Given the face spoof detection classifier working on a single image, a multi-frame fusion scheme is proposed to achieve a more stable face spoof detection performance for a video. The classification results from individual frames are combined by a voting scheme to obtain the spoof detection score for a video. A face video is determined to be genuine if over 50% of its frames are classified as genuine face images. Since some published methods report per video face spoof detection performance using N frames, the multi-frame fusion extension allows us to compare the proposed method's performance with state-of-the-art given the same length of testing videos.

V. Face Spoof Databases

A. Public Domain Face Spoof Databases

To evaluate the effectiveness of spoof detection algorithms, many published papers designed and tested their algorithms on proprietary spoof databases [6], [10], [11], [19]. However, only a few authors have made their face spoof databases publicly available [4], [7]–[9], [44], [45]. In this section, we provide a brief summary of three public-domain face spoof databases: NUAA Photograph Imposter database [8], Idiap REPLAY-ATTACK database [4] and CASIA Face Anti-Spoofing Database [9].[7] There are a couple of other public-domain databases for face spoof detection. For example, the VidTIMIT Audio-Video database (43 subjects) [44] and the DaFEx database (8 subjects) [45] have also been used for the purpose of face spoof detection, but their limited size and spoofing diversity makes them less attractive for use in experimental evaluations.

[7]For simplicity, these three public domain spoof databases are referred to as the NUAA, Idiap and CASIA databases in the rest of this paper.

The NUAA Photograph Imposter database [8], released in 2010, is one of the earliest public-domain spoof databases. It consists of 12,614 images (extracted from 143 videos) of genuine and attack attempts of only 15 subjects. Additionally, only hand-held printed photo attack is included in the NUAA database.

The Idiap REPLAY-ATTACK database [4], released in 2012, consists of 1,300 video recordings of both real-access and attack attempts of 50 different subjects.[8] In the same acquisition condition (controlled and adverse illumination), the face spoof attacks were generated by forging live verification attempts of the same subjects via printed photos, displayed photos/videos on mobile phone's screen, and displayed photos/videos on HD screen.

The CASIA Face Anti-Spoofing Database (FASD) [9], released in 2012, consists of 600 video recordings of genuine and attack attempts of 50 different identities. Although the size of the CASIA database is somewhat smaller than the Idiap database, it contains more diverse samples in terms of the acquisition devices (high resolution Sony NEX-5 camera and low-quality USB camera), face variations (pose and expression variations), and attack attempts (warp photo, cut photo, and HD displayed video).

Table II provides a summary of the above three databases in terms of sample size, acquisition device, attack type, and age, gender and race distributions of subjects. A major drawback of these three spoof databases is that they were all captured by web cameras or high quality digital cameras. There is no public-domain face spoof database using mobile phone cameras as capturing devices. The mobile phone front-facing cameras pose the following additional challenges for face spoof detection: i) They usually have lower resolution, narrow dynamic range, and inaccurate metering and auto-focus capabilities. As a result, videos or images captured by these cameras typically have low quality due to defocus, under or over exposure. Since these image quality degradations appear in both genuine and spoof face images, they will diminish the differences between genuine and spoof face images in terms of facial detail and image distortion. ii) The purpose of building a mobile phone face spoof database is not simply to make the face spoof detection task more difficult, but to better replicate a realistic scenario, e.g., the face unlock applications in Android smartphones using the front-facing cameras.

Another noticeable property of these databases is the standoff distance used in launching the spoof attacks. In the Idiap database, the attacker presented the spoof medium[9] fairly close to the camera (short distance spoofing attack), resulting in a relatively large facial area in the spoof video. In the CASIA database, the spoof attacks were generated with a larger standoff [46] (long distance spoofing attack), resulting in a smaller facial area and lower contrast in the spoof attacks.

[8]The Idiap REPLAY-ATTACK database extended the Idiap PRINT-ATTACK database [7] by augmenting the spoof face samples with a wider range of variability (replayed video and photo attacks).

[9]Three types of medium were used: iPhone 3GS with a 3.5″ screen, iPad with a 9″ screen and 14″ A4 paper. The former two mobile devices were used to replayed videos, and the last medium was used to print photos.

TABLE II
A SUMMARY OF THREE SPOOF FACE DATABASES IN PUBLIC-DOMAIN AND THE MSU MFSD DATABASE

Database	Year of release	# subjects	# videos	Acquisition camera device	Attack type	Subject race	Subject gender	Subject age
NUAA [8]	2010	15	• 24 genuine • 33 spoof	• Web-cam (640 × 480)	• Printed photo	• Asian 100%	• Male 80% • Female 20%	20 to 30 yrs
Idiap REPLAY-ATTACK [4]	2012	50	• 200 genuine • 1,000 spoof	• MacBook 13″ camera (320 × 240)	• Printed photo • Display photo (mobile/HD) • Replayed video (mobile/HD)	• White 76% • Asian 22% • Black 2%	• Male 86% • Female 14%	20 to 40 yrs
CASIA FASD [9]	2012	50	• 150 genuine • 450 spoof	• Low-quality camera (640 × 480) • Normal-quality camera (480 × 640) • Sony NEX-5 camera (1280 × 720)	• Printed photo • Cut photo‡ • Replayed video (HD)	• Asian 100%	• Male 86% • Female 14%	20 to 35 yrs
MSU MFSD	2014	55†	• 110 genuine • 330 spoof	• MacBook Air 13″ camera (640 × 480) • Google Nexus 5 camera (720 × 480)	• Printed photo • Replayed video (mobile/HD)	• White 70% • Asian 28% • Black 2%	• Male 63% • Female 37%	20 to 60 yrs

†Among these 55 subjects, only samples from 35 subjects can be made publicly available based on the subjects' approval.
‡Cut photo attack: a printed photo with the eye regions cut out. The attacker covers his face with the cutout photo and can blink his eyes through the holes.

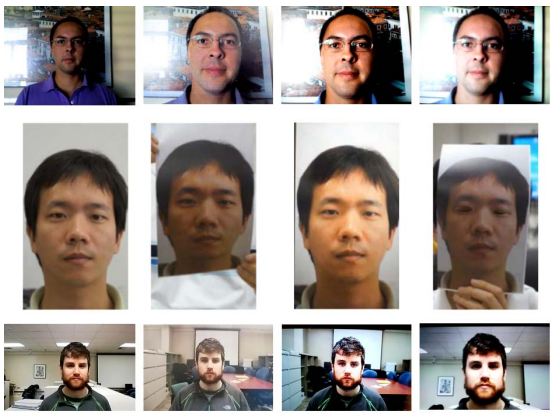


Fig. 6. Typical face samples from the Idiap (first row), CASIA H subset (second row) and MSU (third row) spoofing databases. (a) Genuine face images; (b) Spoof faces generated for printed photo attack; (c) Spoof faces generated by HD tablet screen; (d) Spoof faces generated by mobile phone screen (first and third row) or cut photo (second row).

In this paper, we focus on the short distance spoofing attacks (see Fig. 6) and introduce our own MSU Mobile Face Spoofing Database (MSU MFSD) to facilitate spoof detection research on mobile phone applications.

B. MSU MFSD Database

The MSU MFSD database[10] consists of 440 video clips of photo and video attack attempts of 55 subjects. Two types of cameras were used in collecting this database: i) built-in camera in MacBook Air 13″, referred to as *laptop camera*; ii) front-facing camera in the Google Nexus 5 Android phone, referred to as *Android camera*. Both these devices are the state of the art models. The Nexus 5 phone is also equipped with face unlock and anti-spoofing functionalities using its front-facing camera.

For the laptop camera, videos are captured using the QuickTime framework on the Mac OS X Mavericks platform, with a resolution of 640×480. For the Android camera, videos are captured using the Google built-in camera software on Android 4.4.2, with a resolution of 720 × 480. These video files can be decoded on multiple platforms using the FFmpeg library [47]. The average frame rate is about 30 fps, and the duration of each video is at least 9 seconds (average duration is ∼12 seconds).

1) Genuine Face: The (true) subject presents his face close to the camera, and a genuine face video is recorded using both the Android and laptop cameras. The average standoff distance between the face and the camera is ∼50cm.

2) Spoof Attack - Video Replay: The video of the subject's face is first recorded using a Canon 550D Single-lens reflex (SLR) camera and an iPhone 5S back-facing camera. The Canon 550D SLR camera is used to capture a HD video (1920 × 1088), which is replayed on an iPad Air screen to generate the HD video replay attack. The iPhone 5S is used to capture another HD video (1920 × 1080) that is replayed on the iPhone 5S screen to generate the mobile video replay attack. The average standoff for the HD video replay attack is ∼20cm. The average standoff for the mobile video replay attack is ∼10cm.

3) Spoof Attack - Printed Photo: The Canon 550D camera is also used to capture a HD picture (5184 × 3456) of the subject's face, which is then printed on an A3 paper (11.7″ × 16.5″) using a HP Color Laserjet CP6015xh printer (1200 × 600dpi) to generate a printed photo for attack. The average standoff for the printed photo attack is ∼40cm.

Figure 7 shows example images of genuine and spoof faces of one subject in the MSU database. Compared to other public domain face spoof databases, the MSU database has

[10]For simplicity, the MSU MFSD database is referred to as the MSU database in the rest of this paper.

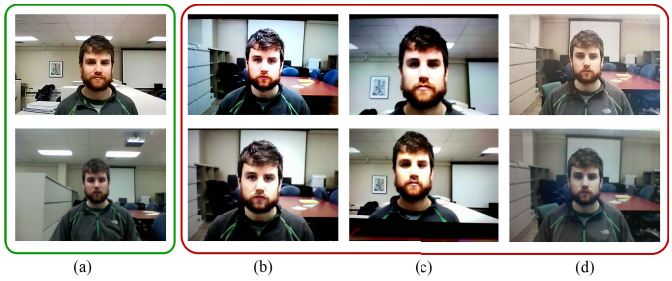


Fig. 7. Example images of genuine and spoof faces of one of the subjects in the MSU MFSD database captured using Google Nexus 5 smart phone camera (top row) and MacBook Air 13″ laptop camera (bottom row). (a) Genuine faces; (b) Spoof faces generated by iPad for video replay attack; (c) Spoof faces generated by iPhone for video replay attack; (d) Spoof faces generated for printed photo attack.

two main desirable properties: i) a mobile phone is used to capture both genuine faces and spoof attacks, simulating the application of mobile phone unlock; and ii) the printed photos used for attacks are generated with a state of the art color printer on larger sized paper. Hence, the printed photos in the MSU database have much better quality than those in the Idiap and CASIA databases. A subset of the MSU database will be made publicly available to interested researcher.[11]

VI. Testing Protocol and Baseline Methods

To evaluate the effectiveness and generalization ability of the proposed face spoof detection methods, we use both intra-database and cross-database protocols. For intra-database testing, we follow the conventional definition of training, developing (if available), and testing sets in the Idiap and CASIA databases. We also define a cross-database protocol to evaluate the generalization ability of spoof detection methods.

A. Intra-Database Testing Protocol

For the Idiap spoof database, we follow the protocols specified in [4] by using all frames in the training and developing sets for training and parameter tuning, respectively. The optimized classifiers are then tested on all frames in the testing set to evaluate the intra-database performance on the Idiap database.

For the CASIA spoof database, the low quality subset (L) and normal quality subset (N) contain only long distance spoof attack samples; only the High quality subset (H) contains short distance spoof attack samples. Therefore, we only test our method under the H protocol specified in [9] for the given training and testing sets.

For the MSU spoof database, there are a total of 55 subjects. We use 35 subjects for the training set and the other 20 for the testing set. The publicly available MSU database (35 subjects) will contain the same testing set (20 subjects) defined here but only a subset of the training set (15 subjects). In this paper, we report the spoof detection performance using both the publicly available MSU database and the complete MSU database.

[11]Following the Institutional Review Board (IRB) rules, only the data of 35 subjects who provided their consent to have their face images released will be made publicly available.

Since there is no developing set in the CASIA and MSU databases, parameter tuning was conducted by cross-validation on the training set. The intra-database performance of the proposed approach on the Idiap and CASIA databases can be directly compared with the state-of-the-art methods.

B. Cross-Database Testing Protocol

The cross-database performance is evaluated by training the spoof detector on database A (e.g., Idiap) and testing it on a different database B (e.g., MSU), and vice versa.

In each cross-database experiment, multi-fold validation is conducted to enumerate different subsets of databases A and B for training and testing, respectively. For example, when using database A for training and database B for testing, we have four possibilities: training on the training set of A and testing on the training/testing set of B; training on the testing set of A and testing on the training/testing set of B. The average performance of this 4-fold evaluation is reported along with variance.

Under the above protocol, a good cross-database performance will provide strong evidence that: i) features are generally invariant to different scenarios (i.e., camera and illuminations), ii) a spoof classifier trained on one scenario is generalizable to the other scenario, and iii) data captured in one scenario can be useful for developing spoof detectors working in the other scenario.

C. Baseline Methods

Most of the published methods (see Table II) have been evaluated on the two public-domain face spoof databases (e.g., Idiap and CASIA) following an intra-database testing protocol. Very few publications have reported the cross-database face spoof detection performance [17].

We have implemented two state of the art methods based on the LBP features [48].[12] These two baseline methods use the same front-end processing described in Section III for face detection and normalization. For each normalized face image, the first method [49] extracts uniform LBP features, including $LBP^{u2}_{8,1}$, $LBP^{u2}_{8,2}$, $LBP^{u2}_{8,3}$, $LBP^{u2}_{8,4}$ and $LBP^{u2}_{16,2}$ channels, constituting a 479-dimensional feature vector.[13] For the second method [50], the 479-dimensional LBP feature vector is extracted after the normalized face image is convolved by a Differences of Gaussian (DoG) filter. This DoG filter was proposed to improve the robustness of LBP features, with $\sigma_1 = 0.5$ and $\sigma_2 = 2$. The resultant LBP features are then fed to train SVM classifiers. These two baseline methods are referred to as LBP+SVM and DoG-LBP+SVM in the experimental results.

[12]While implementations of some published methods based on LBP features are available in public domain, these methods are not applied as baseline here because they do not output the per frame spoof detection performance of the state of the art methods, such as [49], [50].

[13]We have tried extracting holistic LBP features as well as multi-block LBP features (with higher dimensionality). But we found the performance difference between the two to be minor. The authors of [4] also made similar conclusion, namely, enlarging the LBP feature vector does not improve the performance. Since holistic LBP features are computationally more efficient, we choose to extract the 479-dimensional holistic LBP features in our baseline implementation.

TABLE III
NUMBER OF GENUINE AND SPOOF FACES USED IN OUR EXPERIMENT

Database	Training		Developing		Testing	
	Genuine	Spoof	Genuine	Spoof	Genuine	Spoof
Idiap	22,497	69,686	22,498	70,246	29,791	93,686
CASIA (H)	4,579	11,858	–	–	5,603	16,958
MSU	11,567	33,050	–	–	11,178	33,102

TABLE IV
COMPARISON OF INTRA-DATABASE PERFORMANCE ON THE IDIAP, CASIA AND MSU DATABASES

Method	Intra-DB testing on Idiap			Intra-DB testing on CASIA (H protocol)			Intra-DB testing on MSU		
	HTER(%)	TPR@ FAR=0.1	TPR@ FAR=0.01	EER(%)	TPR@ FAR=0.1	TPR@ FAR=0.01	EER(%)	TPR@ FAR=0.1	TPR@ FAR=0.01
LBP-TOP u2 [17]	8.51	≈94.5	≈74.5	13 (75 frms)	≈82	**≈81**	N/A	N/A	N/A
LBP-TOP [16]	7.60	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A
CASIA DoG baseline [9]	N/A	N/A	N/A	26 (30 frms)	≈45	≈14	N/A	N/A	N/A
LBP+SVM baseline	16.1	94.5	57.3	7.5 (30 frms)	**93.3**	29.7	14.7†	69.9†	21.1†
				6.7 (75 frms)	**93.3**	49.0	10.9‡	87.0‡	31.5‡
DoG-LBP+SVM baseline	11.1	92.1	67.0	14.2 (30 frms)	66.7	53.3	23.1†	62.8†	16.4†
				12.7 (75 frms)	84.4	49.7	14.0‡	77.3‡	21.4‡
IDA+SVM (proposed)	7.41	92.2	**87.9**	13.3 (30 frms)	86.7	50	**8.58**†	**92.8**†	**64.0**†
				12.9 (75 frms)	86.7	59.7	**5.82**‡	**94.7**‡	**82.9**‡

†Using the 35 publicly available subjects in the MSU database, 15 for training and 20 for testing.
‡Using all the 55 subjects in the MSU database, 35 for training and 20 for testing.

VII. EXPERIMENTAL RESULTS

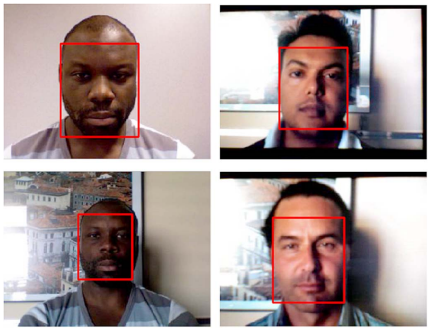
We evaluated three different types of spoof detection feature vectors: LBP features (as used in [49]), DoG-LBP features (as used in [50]), and IDA features defined here. The Idiap REPLAY-ATTACK, CASIA FASD (H protocol), and MSU MFSD databases are used for experiments. The numbers of genuine and spoof faces used in our experiments are detailed in Tab. III. The same classification configuration is adopted for comparing the IDA and other features. When training on a data set with multiple types of spoof attack, the ensemble classifier scheme is used. While training on a data set with only one type of attack, a single classifier scheme is used. Again, for both intra-database and cross-database experiments related to the MSU database, we report the performance on two sets of the MSU MFSD database: the entire dataset with 55 subjects, and a subset with 35 subjects which are publicly available.

A. Intra-Database Spoof Detection

This experiment is to compare the intra-database performance of the proposed method with the baseline methods and state-of-the-art methods on three databases: Idiap, CASIA (H protocol), and the MSU database. The methods for comparison include the LBP-TOP method proposed in [16], the DoG baseline method proposed in [9] and the two baseline methods (LBP+SVM and DoG-LBP+SVM) we implemented. All intra-database results in Table IV are per frame performance unless the number of frames is specified in the parenthesis.

Table IV shows that the proposed IDA+SVM method outperforms the other methods in most intra-database scenarios. The per frame performance on Idiap is even better than LBP-TOP in terms of both HTER and ROC metrics.[14]

[14]The True Positive Rates (TPR) of LBP-TOP in Table IV are estimated from Fig. 3 in [17].



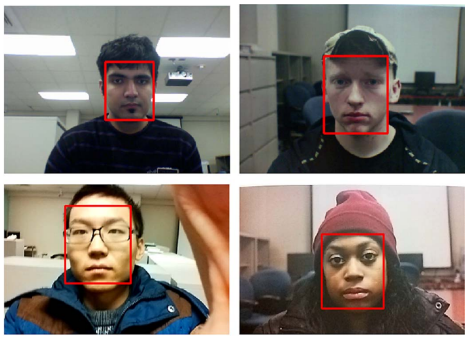
(a) (b)

Fig. 8. Examples of mis-classified face images by the proposed approach in the Idiap intra-database experiment. (a) Genuine face images (dark skin) are misclassified as spoof attacks; (b) Spoof attack face images with relatively small image distortions are misclassified as genuine face images.

In the CASIA (H protocol) testing, performance of the proposed method is close to the two best methods (LBP-TOP and LBP+SVM). In the MSU testing, the proposed method achieves TPR = 94.7% @ FAR = 0.1 and TPR = 82.9% @ FAR = 0.01 when using 35 subjects for training, much better than those of the baseline methods.

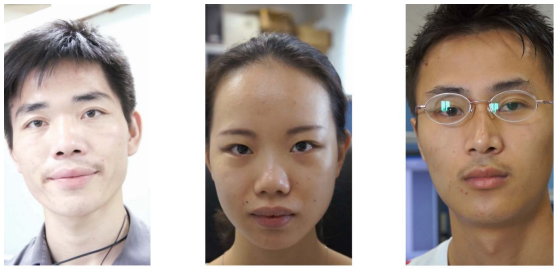
It is well known that texture features, such as LBP and LBP-TOP, have excellent discriminative ability to separate genuine and spoof faces in the same database. In this experiment, we show that the proposed IDA features have similar discriminative ability for intra-database performance. Another baseline is the image quality analysis (IQA) method in [22]. While the IQA method reported a HTER of 15.2% on the Idiap replay-attack database, the proposed method achieved a much lower error rate (HTER = 7.4%) on the same database.

Figures 8 and 9 show some example face images that are incorrectly classified by the proposed IDA+SVM algorithm in the Idiap and MSU intra-database testing, respectively. Each example frame is shown together with its facial



(a) (b)

Fig. 9. False negative and false positive examples in the MSU intra-database experiment. (a) Genuine face images with dark skin (top) and with motion blurriness (bottom); (b) Spoof face images with small image distortion (top) and with dark skin (bottom).



(a) (b) (c)

Fig. 10. Genuine faces with over saturated exposure in the CASIA (H) database, which are mis-classified by the proposed approach in the CASIA (H) intra-database experiment.

cropping region (the red rectangle), in which the IDA feature is extracted. In Fig. 8a, the two false negative samples in the Idiap intra-database testing are from the same dark skinned subject. we believe the lack of dark skinned subjects in the Idiap training set caused these errors. For the two false positive samples in Fig. 8, both are misclassified due to small image distortion in the facial regions.

In Fig. 9, a genuine (top in (a)) and a spoof (bottom in (b)) face from two dark skinned subjects are misclassified. Similar to the Idiap training set, there is no dark skinned subject in the MSU training set, making it difficult to differentiate genuine and spoof faces of dark skinned subjects. The bottom example in Fig. 9 (a) is misclassified due to its motion blurriness. The top example in Fig. 9 (b) (a printed photo attack) is misclassified because of small image distortion.

It is also shown that the proposed method does not perform very well on the CASIA (H) database among the compared methods. The main reasons are as follows: i) the CASIA (H) database has much fewer training samples (20 genuine face videos and 60 spoof face videos) than the other databases (Idiap and MSU) to learn the classifiers; ii) an examination of the misclassified videos shows that some genuine face videos were captured with over saturated exposure, which makes the genuine face images easily confused with their spoof face images (see Fig. 10).

B. Cross-Database Spoof Detection

Besides the intra-database performance, we are more interested in the cross-database performance of different face spoof detection methods. Since the IDA features do not contain any characterization of facial appearance, they are expected to have better cross-database generalization ability than the texture features. In this experiment, we compare the cross-database performance of different features using the three spoof databases (Idiap, CASIA and MSU). Two groups of cross-database performance were evaluated: Idiap-vs-MSU and vice versa and CASIA(H)-vs-MSU and vice versa. As we mentioned in Section II, these three databases are all short-distance spoof databases, but captured by quite different cameras (laptop camera, high quality camera and Android phone camera). Further, illumination conditions in these three databases are different, making it a challenging task to achieve good generalization performance. Experiments are conducted on replay attack samples and printed photo attack samples[15] separately.

Table V compares the Idiap-vs-MSU and CASIA(H)-vs-MSU cross-database performance of different features for replay attack and printed photo attack samples. From these results we make the following observations:

1) IDA Features Outperform LBP and DoG-LBP Features in Cross-Database Testing: In considering different features, the LBP and DoG-LBP results show that they are not able to handle the cross-database challenge. See, for example, the LBP+SVM results in Tables V-c and V-d. On the other hand, the IDA features are more robust, for both replay attack samples and printed attack samples, in almost every cross-database testing scenario. For the replay attack samples, the IDA features achieve nearly perfect separation of genuine and spoof samples in the MSU database with the classifier trained on Idiap database (average TPR $= 90.5\%$ @ FAR $= 0.01$). For the printed attack samples, the IDA features report an average TPR of 31.2% @ FAR $= 0.01$ when trained on Idiap and tested on MSU, which is still much better than the LBP and DoG-LBP features. These results show that the IDA features are, to some extent, able to compensate for different cameras used in the Idiap (laptop camera) and MSU (laptop camera and Android phone camera) databases. It is even possible to utilize a spoof database captured by a laptop camera to design a face spoof detector for mobile phones. Given the significantly better performance of the proposed approach (compared to the IQA method in [22]) with intra-database testing, it is very likely that the proposed approach will outperform the IQA method in [22] under cross-database testing.

2) IDA Features Show Better Cross-Database Performance in the Replay Attack Samples Compared to the Printed Attack Samples: In considering different types of attacks, both the Idiap-vs-MSU and CASIA(H)-vs-MSU experiments show that the IDA features perform better in the replay attack situation. The reason for this is that the printing quality in these three databases is quite different. The MSU printed photos have better quality than the other two databases, making their image distortion more difficult to discern. Furthermore, the specular reflection in some of the printed attack samples is not easy to discern, especially in the Idiap and CASIA databases.

[15] In the CASIA database, the printed photo attack samples are the warp photo and cut photo samples.

TABLE V

COMPARISON OF CROSS-DATABASE PERFORMANCE OF DIFFERENT METHODS. (a) CROSS-DATABASE PERFORMANCE (%) ON REPLAY ATTACK SAMPLES BETWEEN THE IDIAP AND MSU DATABASES. (b) CROSS-DATABASE PERFORMANCE (%) ON PRINTED ATTACK SAMPLES BETWEEN THE IDIAP AND MSU DATABASES. (c) CROSS-DATABASE PERFORMANCE (%) ON REPLAY ATTACK SAMPLES BETWEEN THE CASIA AND MSU DATABASES. (d) CROSS-DATABASE PERFORMANCE (%) ON PRINTED ATTACK SAMPLES BETWEEN THE CASIA AND MSU DATABASES

Method	**Trained on Idiap and tested on MSU**				**Trained on MSU and tested on Idiap**			
	TPR@† FAR=0.1	**TPR@† FAR=0.01**	**TPR@‡ FAR=0.1**	**TPR@‡ FAR=0.01**	**TPR@† FAR=0.1**	**TPR@† FAR=0.01**	**TPR@‡ FAR=0.1**	**TPR@‡ FAR=0.01**
LBP+SVM	34.6±9.8	4.7±2.7	35.7±10.8	4.0±2.5	35.7±8.9	10.4±3.3	42.9±14.2	9.0±2.8
DoG-LBP+SVM	69.0±11.4	30.7±18.9	73.9±16.4	38.3±26.9	54.2±7.8	17.4±7.1	45.1±8.1	10.7±3.8
IDA+SVM (proposed)	**99.6±0.7**	**90.5±5.7**	**99.9±0.1**	**90.4±8.4**	**82.2±8.9**	**47.2±21.2**	**74.9±16.3**	**46.5±20.8**

(a)

Method	**Trained on Idiap and tested on MSU**				**Trained on MSU and tested on Idiap**			
	TPR@† FAR=0.1	**TPR@† FAR=0.01**	**TPR@‡ FAR=0.1**	**TPR@‡ FAR=0.01**	**TPR@† FAR=0.1**	**TPR@† FAR=0.01**	**TPR@‡ FAR=0.1**	**TPR@‡ FAR=0.01**
LBP+SVM	23.4±11.4	9.0±9.7	24.7±10.6	11.5±12.4	12.1±8.6	4.7±4.6	23.8±3.4	8.1±3.3
DoG-LBP+SVM	16.7±1.3	6.6±1.6	15.1±1.7	5.6±1.4	39.2±13.2	20.0±8.9	48.8±21.7	25.8±15.9
IDA+SVM (proposed)	**64.4±1.7**	**31.2±3.7**	**61.1±4.5**	**26.0±2.4**	**72.5±5.3**	**29.4±17.0**	**69.1±2.2**	**38.5±12.8**

(b)

Method	**Trained on CASIA and tested on MSU**				**Trained on MSU and tested on CASIA**			
	TPR@† FAR=0.1	**TPR@† FAR=0.01**	**TPR@‡ FAR=0.1**	**TPR@‡ FAR=0.01**	**TPR@† FAR=0.1**	**TPR@† FAR=0.01**	**TPR@‡ FAR=0.1**	**TPR@‡ FAR=0.01**
LBP+SVM	7.8±3.6	0.2±0.2	6.1±3.6	0	0.8±1.1	0	0.5±0.8	0.1±0.1
DoG-LBP+SVM	14.2±3.8	4.8±5.0	14.7±3.6	6.2±5.5	4.6±0.7	0.2±0.1	5.9±3.6	0.2±0.1
IDA+SVM (proposed)	**67.2±31.8**	**33.8±29.8**	**67.9±31.9**	**43.0±39.8**	**26.9±4.4**	**10.8±3.2**	**28.3±5.8**	**11.9±7.1**

(c)

Method	**Trained on CASIA and tested on MSU**				**Trained on MSU and tested on CASIA**			
	TPR@† FAR=0.1	**TPR@† FAR=0.01**	**TPR@‡ FAR=0.1**	**TPR@‡ FAR=0.01**	**TPR@† FAR=0.1**	**TPR@† FAR=0.01**	**TPR@‡ FAR=0.1**	**TPR@‡ FAR=0.01**
LBP+SVM	2.3±0.5	0	3.0±0.6	1.2±1.4	**6.0±1.0**	**0.5±0.5**	6.9±1.4	0.8±0.5
DoG-LBP+SVM	6.0±1.7	0.4±0.5	10.6±5.2	1.8±2.9	4.1±1.3	0	4.1±1.7	0.1±0.1
IDA+SVM (proposed)	**21.9±4.2**	**1.4±1.1**	**26.9±6.4**	**1.6±1.0**	2.1±0.5	0	**9.1±7.2**	**1.1±1.3**

(d)

†Using all 55 subjects in the MSU database, 35 for training and 20 for testing.
‡Using the 35 publicly available subjects in the MSU database, 15 for training and 20 for testing.

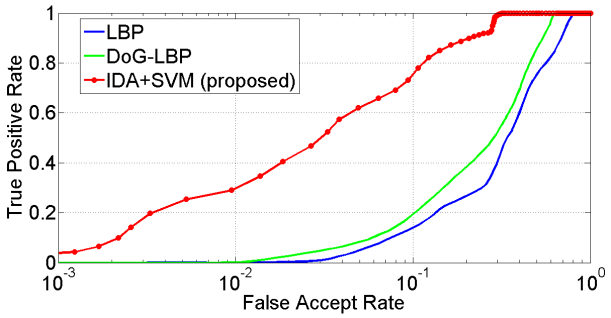
These factors result in different distributions of IDA features for different databases. More robust features for printed photo attack need to be developed.

3) IDA Features Show Better Cross-Database Performance Between Similar Cameras: In considering different cameras, the results show that the IDA features perform better in the Idiap-vs-MSU testing than in the CASIA-vs-MSU testing. Since the camera quality difference between the Idiap (laptop camera) and MSU (laptop and Android phone cameras) databases is smaller than that between CASIA (high quality camera) and MSU databases, it seems that the IDA features can be generalized to cameras with similar quality, but not cameras with distinctively different quality. This might be attributed to the impact of camera quality on the IDA features, e.g., the blurriness measurement. In the CASIA images obtained by their supposedly high quality camera, the blurriness measurement between genuine and spoof faces is not very discriminative compared to the laptop and Android phone cameras used in the MSU database.

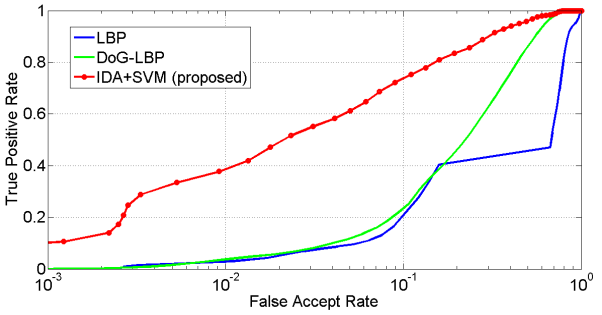
4) IDA Features Achieve the Best Cross-Database Performance With Both Replayed Video and Printed Photo Attacks: The overall cross-database performance for replay and printed photo attack samples, is calculated for different features. For all three features, the same fusion scheme is used. That is, two constituent classifiers are trained using replay attack and printed photo attack videos, respectively. Each testing sample is then passed to these two constituent classifiers and the outputs from the two classifiers are combined in the fusion stage.

In the Idiap-vs-MSU cross-database scenarios, figure 11a shows the overall performance when trained on the Idiap training set and tested on the MSU testing set. Figure 11b shows the overall performance when trained on the MSU training set (using 35 training subjects) and tested on the Idiap testing set. The ROC performance is summarized in Table VI. In these two cross-database scenarios, the IDA features perform the best. TPR of the IDA features @ FAR $= 0.01$ is about 30% better than the DoG-LBP features. Furthermore, the IDA features also show consistent performance in these two bilateral cross-database tests, indicating that the IDA features are more robust across different scenarios than the LBP and DoG-LBP features.

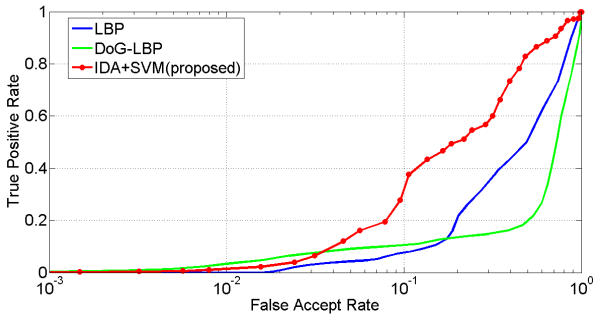
In the CASIA-vs-MSU cross-database scenarios, Figure 11c shows the overall performance when trained on the



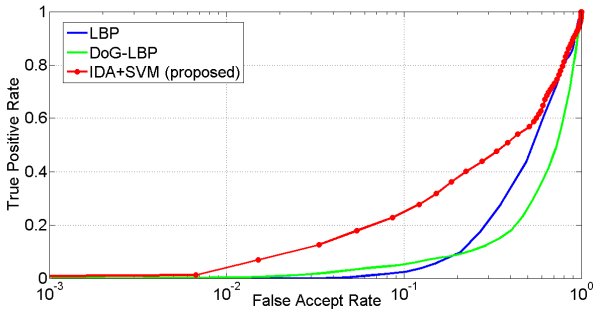
(a)



(b)



(c)



(d)

Fig. 11. Comparison of ROC curves of different features on the Idiap-vs-MSU and CASIA-vs-MSU cross-database testing. (a) Cross-database performance when trained on Idiap and tested on MSU. (b) Cross-database performance when trained on MSU and tested on Idiap. (c) Cross-database performance when trained on CASIA and tested on MSU. (d) Cross-database performance when trained on MSU and tested on CASIA.

CASIA training set and tested on the MSU testing set. Figure 11d shows the overall performance when trained on the MSU training set (using 35 training subjects) and tested on the CASIA testing set. The IDA features again demonstrate great advantage over the other two features.

TABLE VI

INTRA-DATABASE AND CROSS-DATABASE PERFORMANCE (%) OF DIFFERENT METHODS ON BOTH REPLAYED VIDEO AND PRINTED PHOTO ATTACKS

Method	Train	Test	TPR@ FAR=0.1	TPR@ FAR=0.01
LBP+SVM	Idiap	Idiap	94.5	57.3
		MSU	14.1	0
	MSU	MSU	87.0	31.5
		Idiap	20.9	2.9
DoG-LBP +SVM	Idiap	Idiap	92.1	67.0
		MSU	19.5	0.2
	MSU	MSU	77.3	21.4
		Idiap	23.6	3.8
IDA+SVM (proposed)	Idiap	Idiap	92.2	**87.9**
		MSU	**75.5**	**29.8**
	MSU	MSU	**94.7**	**82.9**
		Idiap	**73.7**	**38.6**

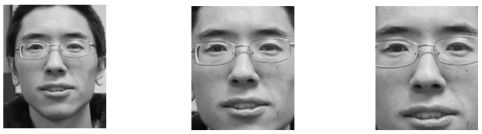


Fig. 12. Three different face cropping sizes chosen for cross-database performance comparison in our experiment.

We have also evaluated the effect of different face cropping sizes on the cross-database spoof detection performance. Three different face cropping sizes are chosen to extract the IDA features, with interpupillary distances (IPD) of 44, 60 and 70 pixels (see Fig. 12). As shown in Fig. 12, the bigger the IPD, the smaller the contextual information included in feature extraction. We then perform cross-database testing on the MSU database by training the proposed approach on Idiap and CASIA databases, respectively. The cross-DB ROC performance is shown in Fig. 13.

We can notice that: i) the face cropping with small amount of contextual information leads to the best performance; ii) small performance degradation is observed when no context information is included in the face images (e.g., IPD $=$ 70), and iii) large performance degradation is observed when more context information is included in face images (e.g., IPD $=$ 44). The reason is that small context information is helpful in differentiating depth in real and spoof faces. However, the context region also has different reflectance property than the facial region, which is irrelevant to face spoof detection. As a result, large context region degrades the cross-DB performance.

Figure 14 shows some example face images in the MSU database that are correctly classified by the proposed IDA+SVM algorithm in the cross-database testing, indicating that the IDA features are capable of differentiating genuine and spoof faces in different acquisition conditions.

In the cross-database testing, we find that most errors are caused when the testing sample has unusual image quality, color distribution, or acquisition condition that is not represented in the training samples. Figure 15 shows some

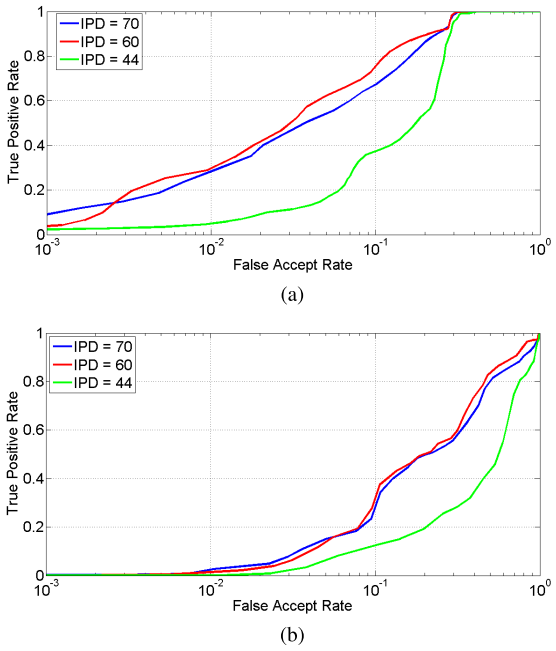


Fig. 13. Comparison of ROC curves of different face cropping sizes achieved on the cross-database testing. (a) Cross-database performance on MSU DB (trained on Idiap DB) with different face cropping sizes. (b) Cross-database performance on MSU DB (trained on CASIA(H) DB) with different face cropping sizes.

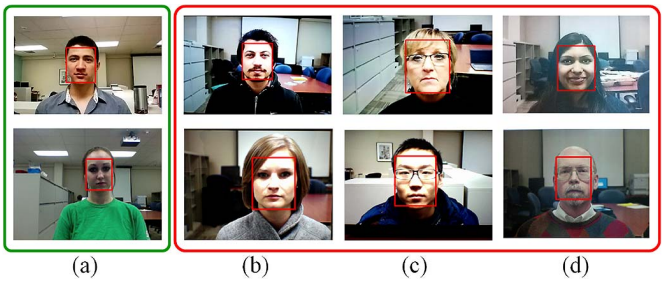


Fig. 14. Examples of correct face spoof detection by the proposed approach on the MSU database with cross-database testing protocol (classifier trained on the Idiap database). (a) Genuine face images; (b) Video replay spoof attacks by an iPad; (c) Video replay spoof attacks by an iPhone; (d) Printed photo spoof attacks with A3 paper.

example face images that are incorrectly classified by the proposed IDA+SVM algorithm in the cross-database testing. Figures 15 (a,b) show four false negative examples. The first MSU example (upper left: captured by laptop camera) is misclassified due to its large standoff distance, which is unusual for the Idiap training samples. But this example can be correctly classified in the MSU intra-database experiment, indicating that the MSU classifier can tolerate such a standoff because there are similar samples in the MSU training set. The second MSU example (upper right: captured by Android camera) is misclassified because of its apparent motion blurriness. This is caused by the severe hand-held motion in the Android phone captured samples. Note that the Idiap genuine samples were all captured by a stationary camera.

The other two examples in Fig. 15 (b) are misclassified because of the subject's apparent dark skin (lower left) and

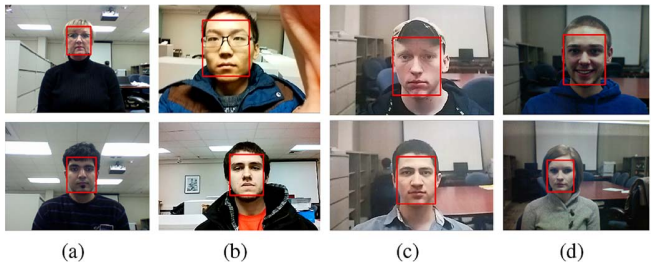


Fig. 15. Examples of false negative and false positive samples in the cross-database experiment (trained on Idiap and tested on MSU). (a) Misclassified genuine samples due to large standoff and low contrast; (b) Misclassified genuine samples due to motion blur and over saturation; (c) Spoof samples captured by the Android camera that were classified as genuine; (d) Spoof samples captured by the laptop camera that were classified as genuine.

over-saturated white illumination (lower right). Dark skin appearance has different chromatic properties compared to white skin, especially in the skewness of hue histogram. Since there is no dark skinned subject in the Idiap training set, the classifier trained on the Idiap dataset is not able to correctly classify dark skinned subjects in the MSU testing set. Over-saturated white illumination on a genuine face also diminishes the difference between genuine face and spoof face in the proposed feature space. For example, the specular reflection extraction method will incorrectly extract the whole face region as specular component, resulting in a high specular pixel ratio r. Since there is no genuine sample in the Idiap training set under over-saturated illumination, these errors are inevitable in cross-database testing. The errors caused by insufficient dark skinned subjects in the training set suggest that race balance in spoof database should be considered.

Four false positive examples in the cross-database testing are also shown in Figs. 15 (c,d). All of them are printed photo attack samples. As mentioned earlier, printed photo attack samples in the MSU database have much better quality than those in the Idiap database, diminishing the genuine-to-spoof difference in the mobile camera capture. Figures 15 (c,d) show that the color distribution of these four example images is very close to the genuine faces and almost no specular reflection can be observed. Therefore, image distortion in these spoof samples is quite small.

C. Intra-Database vs. Cross-Database

Given the above experimental results, we can further compare the cross-database performance with intra-database performance of the same method. Table VI shows that the proposed method performs much better than the LBP+SVM and DoG-LBP+SVM methods in its generalization ability.

When trained on the Idiap database, the proposed method not only achieves nearly the best performance on the Idiap testing set, but also achieves a more stable performance on the MSU testing set (TPR = 75.5% @ FAR = 0.1, TPR = 29.8% @ FAR = 0.01) than the LBP+SVM and DoG-LBP+SVM methods. When trained on the MSU database, the proposed method performs the best in both intra-database and cross-database testings, much better than the two baseline methods. This comparison reveals that

TABLE VII
COMPUTATIONAL REQUIREMENT OF THE IDA+SVM ALGORITHM FOR AN ANDROID PHONE VIDEO WITH 300 FRAMES

Operation	Face normalization	IDA feature extraction	Classification	Total
Avg. time per frame (second)	0.12	0.12	0.02	0.26

the baseline (texture based) face spoof detection methods get over-trained on a specific spoof database, causing their performance to degrade significantly on a different testing database. The proposed IDA features appear to be more robust than texture features.

Recall that in Table V, the proposed method has even better cross-database performance in two separate attack groups than in the overall testing (see Table VI). This suggests that most errors of the proposed method are caused by the fusion scheme.

D. Computational Time Requirement

The computation cost of most of the published methods on face spoof detection is unknown. In [12], the HOOF+LDA (NN) approach required about 0.9 sec. per frame. Further, this method needed 230 frames to compute the HOOF feature and can not work with only a single image or frame.

Table VII lists the computational requirement of our IDA+SVM algorithm on a 720×480 Android phone video. Not including the time needed for face detection, the total time consumption for face normalization, feature extraction, and classification (including the two constituent classifiers) is about 0.26 sec. per frame on a testing platform with Intel$^{(R)}$ Core$^{(TM)}$ i7-4700MQ CPU @ 2.40 GHz and 8GB RAM. Our method can output a spoof detection result based on either a single image or multiple video frames. Currently, the proposed approach is implemented in MATLAB, likely allowing for further optimizations.

VIII. CONCLUSIONS AND FUTURE WORK

In this paper, we address the problem of face spoof detection, particularly in a cross-database scenario. While most of the published methods use motion or texture based features, we propose to perform face spoof detection based on Image Distortion Analysis (IDA). Four types of IDA features (specular reflection, blurriness, color moments, and color diversity) have been designed to capture the image distortion in the spoof face images. The four different features are concatenated together, resulting in a 121-dimensional IDA feature vector. An ensemble classifier consisting of two constituent SVM classifiers trained for different spoof attacks is used for the classification of genuine and spoof faces. We have also collected a face spoof database, called MSU MFSD, using two mobile devices (Android Nexus 5, and MacBook Air $13''$). To our knowledge, this is the first mobile spoof face database. A subset of this database, consisting of 35 subjects, will be made publicly available (http://biometrics.cse.msu.edu/pubs/databases.html).

Evaluations on three face spoof databases (Idiap REPLAY-ATTACK, CASIA FASD, and MSU MFSD) show that the proposed approach performs better than the state-of-the-art methods in intra-database testing scenario and significantly outperforms the baseline methods in cross-database scenario.

Our suggestions for future work on face spoof detection include: (i) understand the characteristics and requirements of the use case scenarios for face spoof detection, (ii) collect a large and representative database that considers the user demographics (age, gender, and race) and ambient illumination in the use case scenario of interest, (iii) develop robust, effective, and efficient features (e.g., through feature transformations [51]) for the selected use case scenario, and (iv) consider user-specific training for face spoof detection.

ACKNOWLEDGMENT

The authors would like to thank the Idiap and CASIA institutes for sharing their face spoof databases. This manuscript benefited from the valuable comments provided by the editors and reviewers.

REFERENCES

[1] A. Rattani, N. Poh, and A. Ross, "Analysis of user-specific score characteristics for spoof biometric attacks," in *Proc. IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recognit. Workshops (CVPRW)*, Jun. 2012, pp. 124–129.

[2] Z. Wu, N. Evans, T. Kinnunen, J. Yamagishi, F. Alegre, and H. Li, "Spoofing and countermeasures for speaker verification: A survey," *Speech Commun.*, vol. 66, pp. 130–153, Feb. 2015.

[3] L. Best-Rowden, H. Han, C. Otto, B. F. Klare, and A. K. Jain, "Unconstrained face recognition: Identifying a person of interest from a media collection," *IEEE Trans. Inf. Forensics Security*, vol. 9, no. 12, pp. 2144–2157, Dec. 2014.

[4] I. Chingovska, A. Anjos, and S. Marcel, "On the effectiveness of local binary patterns in face anti-spoofing," in *Proc. IEEE BIOSIG*, Sep. 2012, pp. 1–7.

[5] N. Erdogmus and S. Marcel, "Spoofing in 2D face recognition with 3D masks and anti-spoofing with Kinect," in *Proc. IEEE BTAS*, Sep./Oct. 2013, pp. 1–6.

[6] Z. Zhang, D. Yi, Z. Lei, and S. Z. Li, "Face liveness detection by learning multispectral reflectance distributions," in *Proc. FG*, Mar. 2011, pp. 436–441.

[7] A. Anjos and S. Marcel, "Counter-measures to photo attacks in face recognition: A public database and a baseline," in *Proc. IJCB*, Oct. 2011, pp. 1–7.

[8] 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. ECCV*, Sep. 2010, pp. 504–517.

[9] Z. Zhang, J. Yan, S. Liu, Z. Lei, D. Yi, and S. Z. Li, "A face antispoofing database with diverse attacks," in *Proc. ICB*, Mar./Apr. 2012, pp. 26–31.

[10] L. Sun, G. Pan, Z. Wu, and S. Lao, "Blinking-based live face detection using conditional random fields," in *Proc. AIB*, 2007, pp. 252–260.

[11] W. Bao, H. Li, N. Li, and W. Jiang, "A liveness detection method for face recognition based on optical flow field," in *Proc. IASP*, Apr. 2009, pp. 233–236.

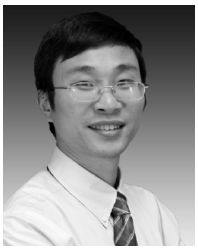
[12] S. Bharadwaj, T. I. Dhamecha, M. Vatsa, and R. Singh, "Computationally efficient face spoofing detection with motion magnification," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. Workshops (CVPRW)*, Jun. 2013, pp. 105–110.

[13] 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.

[14] *The TABULA RASA Project*. [Online]. Available: http://www.tabularasa-euproject.org/, acccessed Sep. 2014.

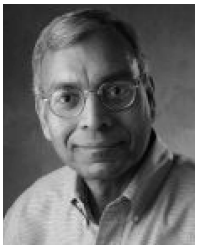
[15] K. Kollreider, H. Fronthaler, M. I. Faraj, and J. Bigun, "Real-time face detection and motion analysis with application in 'liveness' assessment," *IEEE Trans. Inf. Forensics Security*, vol. 2, no. 3, pp. 548–558, Sep. 2007.

[16] T. de Freitas Pereira, A. Anjos, J. M. De Martino, and S. Marcel, "*LBP–TOP* based countermeasure against face spoofing attacks," in *Proc. ACCV Workshops*, 2012, pp. 121–132.
[17] 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. ICB*, Jun. 2013, pp. 1–8.
[18] J. Yang, Z. Lei, S. Liao, and S. Z. Li, "Face liveness detection with component dependent descriptor," in *Proc. IJCB*, Jun. 2013, pp. 1–6.
[19] 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. ICB*, Jun. 2013, pp. 1–6.
[20] J. Komulainen, A. Hadid, and M. Pietikäinen, "Context based face anti-spoofing," in *Proc. BTAS*, Sep./Oct. 2013, pp. 1–8.
[21] G. Chetty, "Biometric liveness checking using multimodal fuzzy fusion," in *Proc. IEEE FUZZ*, Jul. 2010, pp. 1–8.
[22] J. Galbally, S. Marcel, and J. Fierrez, "Image quality assessment for fake biometric detection: Application to iris, fingerprint, and face recognition," *IEEE Trans. Image Process.*, vol. 23, no. 2, pp. 710–724, Feb. 2014.
[23] H.-Y. Wu, M. Rubinstein, E. Shih, J. Guttag, F. Durand, and W. Freeman, "Eulerian video magnification for revealing subtle changes in the world," *ACM Trans. Graph.*, vol. 31, no. 4, Jul. 2012, Art. ID 65.
[24] S. A. Shafer, "Using color to separate reflection components," *Color Res. Appl.*, vol. 10, no. 4, pp. 210–218, 1985.
[25] O. Bimber and D. Iwai, "Superimposing dynamic range," in *Proc. ACM SIGGRAPH Asia*, 2008, pp. 1–8, no. 150.
[26] PittPatt Software Developer Kit. *Pittsburgh Pattern Recognition PittPatt*. [Online]. Available: http://www.pittpatt.com/, accessed Jan. 2011.
[27] Q. Yang, S. Wang, and N. Ahuja, "Real-time specular highlight removal using bilateral filtering," in *Proc. ECCV*, 2010, pp. 87–100.
[28] V. Christlein, C. Riess, E. Angelopoulou, G. Evangelopoulos, and I. Kakadiaris, "The impact of specular highlights on 3D-2D face recognition," *Proc. SPIE*, vol. 8712, p. 87120T, May 2013.
[29] R. T. Tan and K. Ikeuchi, "Separating reflection components of textured surfaces using a single image," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 27, no. 2, pp. 178–193, Feb. 2005.
[30] J.-F. Lalonde, A. A. Efros, and S. G. Narasimhan, "Estimating the natural illumination conditions from a single outdoor image," *Int. J. Comput. Vis.*, vol. 98, no. 2, pp. 123–145, Jun. 2011.
[31] H. Han, S. Shan, X. Chen, S. Lao, and W. Gao, "Separability oriented preprocessing for illumination-insensitive face recognition," in *Proc. ECCV*, 2012, pp. 307–320.
[32] X. Gao, T.-T. Ng, B. Qiu, and S.-F. Chang, "Single-view recaptured image detection based on physics-based features," in *Proc. ICME*, Jul. 2010, pp. 1469–1474.
[33] F. Crete, T. Dolmiere, P. Ladret, and M. Nicolas, "The blur effect: Perception and estimation with a new no-reference perceptual blur metric," *Proc. SPIE*, vol. 6492, p. 64920I, Feb. 2007.
[34] P. Marziliano, F. Dufaux, S. Winkler, and T. Ebrahimi, "A no-reference perceptual blur metric," in *Proc. ICIP*, vol. 3. 2002, pp. III-57–III-60.
[35] Y. Chen, Z. Li, M. Li, and W.-Y. Ma, "Automatic classification of photographs and graphics," in *Proc. ICME*, Jul. 2006, pp. 973–976.
[36] B. E. Boser, I. M. Guyon, and V. N. Vapnik, "A training algorithm for optimal margin classifiers," in *Proc. 5th ACM Workshop Comput. Learn. Theory*, 1992, pp. 144–152.
[37] A. Bashashati, M. Fatourechi, R. K. Ward, and G. E. Birch, "A survey of signal processing algorithms in brain–computer interfaces based on electrical brain signals," *J. Neural Eng.*, vol. 4, no. 2, pp. R32–R57, Mar. 2007.
[38] C. Hou, F. Nie, C. Zhang, D. Yi, and Y. Wu, "Multiple rank multi-linear SVM for matrix data classification," *Pattern Recognit.*, vol. 47, no. 1, pp. 454–469, Jan. 2014.
[39] Y. Lin *et al.*, "Large-scale image classification: Fast feature extraction and SVM training," in *Proc. IEEE CVPR*, Jun. 2011, pp. 1689–1696.
[40] C.-C. Chang and C.-J. Lin, "LIBSVM: A library for support vector machines," *ACM Trans. Intell. Syst. Technol.*, vol. 2, no. 3, pp. 27:1–27:27, May 2011.
[41] 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, Aug. 2008.
[42] F. Nie, Y. Huang, X. Wang, and H. Huang, "New primal SVM solver with linear computational cost for big data classifications," in *Proc. ICML*, 2014, pp. 1–9.
[43] A. Jain, K. Nandakumar, and A. Ross, "Score normalization in multimodal biometric systems," *Pattern Recognit.*, vol. 38, no. 12, pp. 2270–2285, Dec. 2005.
[44] C. Sanderson, *Biometric Person Recognition: Face, Speech and Fusion*. Saarbrücken, Germany: VDM-Verlag, 2008.
[45] A. Battocchi and F. Pianesi, "DaFEx: Un database di espressioni facciali dinamiche," in *Proc. SLI-GSCP Workshop*, 2004, pp. 311–324.
[46] T. de Freitas Pereira *et al.*, "Face liveness detection using dynamic texture," *EURASIP J. Image Video Process.*, vol. 2014, no. 1, p. 2, Jan. 2014.
[47] *The FFmpeg Multimedia Framework*. [Online]. Available: http://www.ffmpeg.org/, accessed Jan. 2014.
[48] T. Ojala, M. Pietikäinen, and T. Mäenpää, "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.
[49] J. Määttä, A. Hadid, and M. Pietikäinen, "Face spoofing detection from single images using micro-texture analysis," in *Proc. IJCB*, Oct. 2011, pp. 1–7.
[50] N. Kose and J. Dugelay, "Classification of captured and recaptured images to detect photograph spoofing," in *Proc. ICIEV*, May 2012, pp. 1027–1032.
[51] C. Zhang, F. Nie, and S. Xiang, "A general kernelization framework for learning algorithms based on kernel PCA," *Neurocomputing*, vol. 73, nos. 4–6, pp. 959–967, Jan. 2010.



Di Wen (M'15) received the B.S. degree in information and electronic engineering from Zhejiang University, Hangzhou, China, in 1999, and the Ph.D. degree in electronics engineering from Tsinghua University, Beijing, China, in 2007. From 2013 to 2014, he was a Research Associate with the Department of Computer Science, Michigan State University, East Lansing, MI, USA. He is currently a Research Associate with the Department of Biomedical Engineering, Case Western Reserve University, Cleveland, OH, USA. He is involved in particular on face recognition and retrieval in video surveillance, object detection in video surveillance, facial demographic recognition, and medical image analysis. His research interests include computer vision, pattern recognition, and image processing.

Hu Han (M'13) received the B.S. degree from Shandong University, Jinan, China, and the Ph.D. degree from the Institute of Computing Technology, Chinese Academy of Sciences, Beijing, China, in 2005 and 2011, respectively, all in computer science. He is currently a Research Associate with the Department of Computer Science and Engineering, Michigan State University, East Lansing, MI, USA. His research interests include computer vision, pattern recognition, and image processing, with applications to biometrics, forensics, law enforcement, and security systems.



Anil K. Jain (LF'14) is currently a University Distinguished Professor with the Department of Computer Science and Engineering, Michigan State University, East Lansing, MI, USA. He has coauthored a number of books, including *Handbook of Fingerprint Recognition* (2009), *Handbook of Biometrics* (2007), *Handbook of Multibiometrics* (2006), *Handbook of Face Recognition* (2005), *Biometrics: Personal Identification in Networked Society* (1999), and *Algorithms for Clustering Data* (1988). His research interests include pattern recognition and biometric authentication. He is a fellow of the American Association for the Advancement of Science, the Association for Computing Machinery, the International Association for Pattern Recognition (IAPR), and the International Society for Optics and Photonics. He served as a member of the Defense Science Board and the National Academies Committees on Whither Biometrics and Improvised Explosive Devices. He was a recipient of the 1996 IEEE TRANSACTIONS ON NEURAL NETWORKS Outstanding Paper Award and the Pattern Recognition Society Best Paper Awards in 1987, 1991, and 2005. He has received the Fulbright Award, the Guggenheim Award, the Humboldt Award, the IEEE Computer Society Technical Achievement Award, the IEEE Wallace McDowell Award, the ICDM Research Contributions Award, and the IAPR King-Sun Fu Award. He served as the Editor-in-Chief of the IEEE TRANSACTIONS ON PATTERN ANALYSIS AND MACHINE INTELLIGENCE (1991–1994).