Face Liveness Detection Based on Perceptual Image Quality Assessment Features with Multi-scale Analysis

Chun-Hsiao Yeh Herng-Hua Chang
Computational Biomedical Engineering Lab
Department of Engineering Science and Ocean Engineering
National Taiwan University, Taiwan
{r04525061, herbertchang}@ntu.edu.tw

Abstract

Vulnerability of recognition systems to spoofing attacks (presentation attacks) is still an open security issue in the biometrics domain. Among all biometric traits, face is exposed to the most serious threat since it is particularly easy to access and reproduce. In this paper, an effective approach against face spoofing attacks based on perceptual image quality assessment features with multi-scale analysis is presented. First, we demonstrate that the recently proposed blind image quality evaluator (BIQE) is effective in detecting spoofing attacks. Next, we combine the BIQE with an image quality assessment model called effective pixel similarity deviation (EPSD), which we propose to obtain the standard deviation of the gradient magnitude similarity map by selecting effective pixels in the image. A total number of 21 features acquired from the BIQE and EPSD constitute the multi-scale descriptor for classification. Extensive experiments based on both intra-dataset and cross-dataset protocols were performed using three existing benchmarks, namely, Replay-Attack, CASIA, and UVAD. The proposed algorithm demonstrated its superiority in detecting face spoofing attacks over many state of the art methods. We believe that the incorporation of the image quality assessment knowledge into face liveness detection is promising to improve the overall accuracy.

1. Introduction

Liveness detection has become an important issue in the biometric domain for protecting authentication systems. This is because it is relatively easier and simpler to create threatening duplicates of many existing biometrics thanks to modern technologies. Those biometric systems will be invaded if liveness protection mechanisms are absent. For example, fingerprint scanners can be spoofed by common household appliances, e.g., inkjet printers and iris identification systems can be hacked by high-resolution printed eye images. Further examples are the attacks of face recognition systems [1, 2], such as printed face photos, recorded videos, and 3D mask models with different expressions [3]. In order to address these attacks, numerous anti-spoofing security systems and devices have been proposed in the literature.

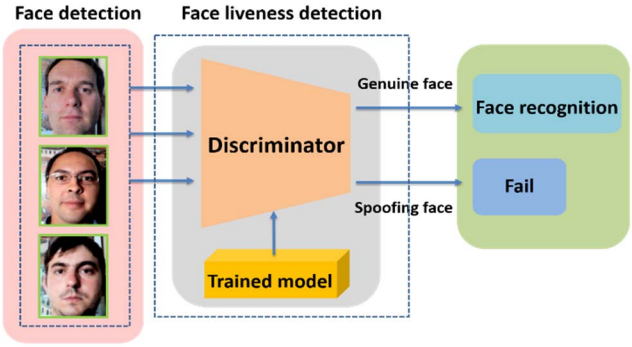

Figure 1: Overview of a contemporary face biometric system: face detection, face liveness detection, and face recognition.

Among the existing anti-spoofing biometric techniques, face recognition has been one of the most studied and widely adopted modalities. One of the reasons is that an abundance of facial traits is available on public websites and social networks such as Facebook, Youtube, and Instagram. Hackers are able to collect facial samples of a specific person from these sources, which can be used to subvert a face recognition system. One easy way to intrude security systems is to present a counterfeit of a genuine enrollee by either using a 2D printed face photo image or displaying the face with an electronic screen. As such, numerous countermeasures against these attacks on face recognition systems are proposed as described in the survey literature [4, 5]. The evolution of the established standards demonstrates the importance to successfully detect spoofing attacks in real-life scenarios.

As depicted in Figure 1, a contemporary face biometric system consists of three major components: face detection, face liveness detection, and face recognition. In this paper, we focus on face liveness detection for anti-spoofing. A new face liveness detection algorithm for detecting video-

978-1-5386-4886-5/18/$31.00 ©2018 IEEE
DOI 10.1109/WACV.2018.00012

based spoofing attacks is proposed based on the assumption that genuine and spoofing face samples (images) contain completely different perceptual quality traits. To the best of our knowledge, this is the first attempt to deal with video-based spoofing attacks in light of image quality assessment features for face liveness detection.

The philosophy underlying our proposed algorithm is based on the fact that synthetic biometric samples usually contain certain artifacts, which are generated during the recapturing and manufacturing process. We address two significant artifacts:

- *Luminance artifact:* Recaptured videos tend to show different distributions of illumination from the original videos. This is often caused by the incomplete color reproduction ability of the display media.
- *Distortion artifact:* During the recapturing process, the detailed information of facial samples is reduced and the blurring effect appears due to re-quantization of the original sample.

The observation of these two facts motivates us to take advantage of the deterioration measure of the facial information in videos. For the luminance artifact, we introduce luminance quality-aware features to characterize structure and contrast distortion. For the distortion artifact, we propose gradient magnitude similarity features, which are sensitive to artifacts generated by compression and blurriness. The proposed features are well able to distinguish recaptured face videos from original face videos.

Through the analysis of quality traits between genuine and spoofing face images, we will design a multi-scale feature descriptor, into which combines various image quality assessments, for detecting face spoofing attacks. The main contributions of this work are summarized as follows:

(i) A pioneering attempt on face liveness detection using image quality assessment features is made to construct a multi-scale descriptor.

(ii) A new image quality assessment method called the effective pixel similarity deviation (EPSD) is introduced based on the improvement of the gradient magnitude similarity deviation (GMSD) [6].

(iii) Three existing datasets of Replay-Attack [7], CASIA [8], and UVAD [9] were greatly utilized for the performance demonstration of the proposed algorithm.

(iv) Extensive experiments with both intra-dataset and cross-dataset protocols in fair comparison with the state of the art methods are conducted to evaluate the detection performance.

(v) A detailed study of the video-based spoofing attacks is administered that concludes the importance of new anti-spoofing techniques to safety.

The remainder of this paper is organized as follows: In Section 2, we give a brief review of face liveness detection related work. We then describe the proposed methodology in Section 3. The experimental results are presented and discussed in Section 4. Finally, the conclusion is drawn in Section 5.

2. Related Work

A variety of anti-spoofing countermeasures have been proposed in order to overcome the invasion of face spoofing attacks. While it is possible to exploit different visual cues for face-spoofing detection, the existing anti-spoofing techniques can be roughly categorized into three groups: methods based on frequency analysis, texture-based analysis, and other cues-based approaches. In what follows, we review the literature based on the classified face spoofing methods, since such methods are non-intrusive and do not require extra devices, which are preferable in practice. For further review, readers please refer to the work of Galbally et al. [10] and Marcel et al. [11].

2.1. Frequency-based approaches

Pinto et al. [12] introduced a technique using visual rhythm analysis for detecting video attacks. According to the authors' perspective, the recaptured video (spoofing attack) included random noise signature. As such, a low-pass filter was conducted to remove the noise signal and the temporal information of the video was captured by visual rhythm techniques. Steiner et al. [13] presented an approach to ensure the authenticity of a face by enhancing existing face verification solutions using the multispectral short wave infrared (SWIR) method.

2.2. Texture-based approaches

Määttä et al. [14] explored image texture features using the Local Binary Pattern (LBP) for detecting spoofing attacks. This approach focused on how testing images were influenced by the luminance and quality. According to the authors' evaluation, the algorithm achieved the area under curve (AUC) of 99% on the NUAA database [15]. Tirunagari et al. [16] applied the dynamic mode decomposition (DMD) as a general purpose to capture liveness cues and proposed a classification pipeline consisting of DMD, LBP, and support vector machines (SVMs) with a histogram intersection kernel to determine genuineness.

Arashloo et al. [17] presented a countermeasure to detect spoofing attacks in various imaging conditions. They combined binarized statistical image features on three orthogonal planes (MBSIF-TOP) with a blur-tolerant descriptor, which is executed via a kernel fusion approach based on the kernel discriminant analysis (KDA) technique. Boulkenafet et al. [18] exploited the joint color-texture information from the luminance and chrominance channels by extracting low-level features in different color spaces. The luminance information of face images was employed

for discriminating spoofing face images from genuine face images. Chan et al. [19] proposed a face liveness detection method against 2D spoofing attacks using four texture features and 2D structure descriptors with low computational complexity.

2.3. Other cue-based approaches

Yang et al. [20] proposed a person-specific face anti-spoofing approach using a classifier specifically trained for each subject, which reduced the interference among subjects. Considering the scarcity for training, they presented a subject domain adaptation method to synthesize virtual features. Wen et al. [21] combined different types of features (specular reflection, blurriness, chromatic moment, and color diversity) to construct the image distortion analysis (IDA) feature vector for detecting spoofing attacks.

Patel et al. [22] developed a face spoofing detection system for an Android smartphone by analyzing image distortion based on intensity channels (R, G, B, and grayscale) and image regions (entire face, detected face, and facial component between the nose and chin). Manjani et al. [23] proposed a multi-level deep dictionary learning-based algorithm that formulated layer-by-layer training to learn deep dictionaries followed by an SVM, which discerned different kinds of attacks, especially the silicone mask attack. Recently, deep learning techniques [24, 25] have been introduced to achieve better detection accuracy.

3. Proposed Methodology

In contrast to the methods described in Section 2, we present a new anti-spoofing solution based on multi-scale image quality feature models. Our framework consists of five major procedures: (i) image preprocessing, (ii) luminance quality-aware feature, (iii) gradient magnitude similarity feature, (iv) feature descriptor, and (v) classification as illustrated in Figure 2.

3.1. Luminance quality-aware feature

The mean subtracted and contract normalized (MSCN) coefficients [26] have been widely employed in image quality assessment (IQA). This is mainly because the distortion in image changes the statistics of the MSCN coefficients, which is sensitive to the assessment of image quality. Stealing from the concept of the MSCN strategy, we develop modified MSCN coefficients for face liveness detection. A rescaled image $\hat{I}(i,j)$ of the nose RGB intensity subimage $I(i,j)$ is first computed using

$$\hat{I}(i,j) = \frac{I(i,j) - \mu_I(i,j)}{\sigma_I(i,j) + 1}, \tag{1}$$

where $i \in \{1, 2 \dots M\}$ and $j \in \{1, 2 \dots N\}$ are the coordinate indices with M and N the image height and width,

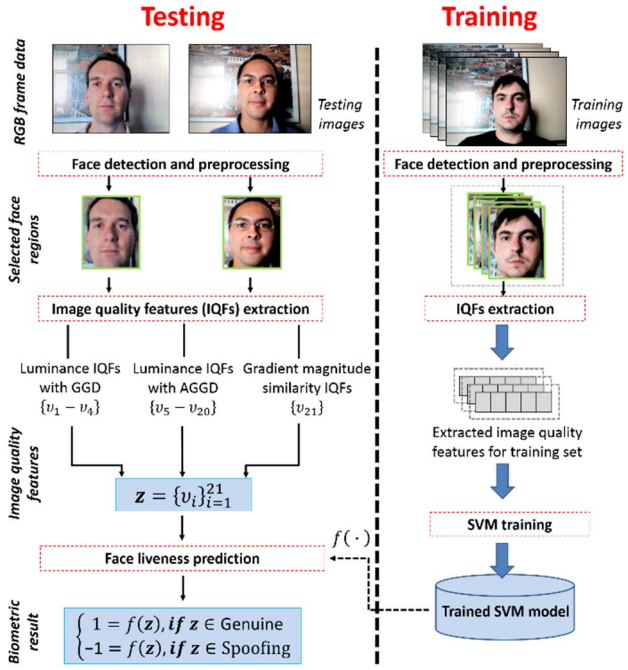

Figure 2: Workflow of the major procedures in our proposed face liveness detection framework.

respectively, μ_I is the local mean:

$$\mu_I(i,j) = \sum_{k=-K}^{K} \sum_{l=-L}^{L} \omega_{k,l}\, I(i+k, j+l)\,, \tag{2}$$

and σ_I is the local standard deviation

$$\sigma_I(i,j) = \sqrt{\sum_{k=-K}^{K} \sum_{l=-L}^{L} \omega_{k,l}\, [I(i+k, j+l) - \mu_I(i,j)]^2}\,, \tag{3}$$

where $\omega = \{\omega_{k,l} | k = -K, \dots, K, l = -L, \dots, L\}$ defines the unit-volume Gaussian window.

Our assumption is that the statistical properties of the MSCN coefficients are easily changed by the distortion, which is highly possible to affect the perceptual image quality. Examples of such quality measures are the level of blurriness and the Laplacian appearance in an image. The generalized Gaussian distribution (GGD) function with zero mean is appropriate to catch the behavior of the MSCN coefficients, which is defined as

$$f(x; v, \sigma^2) = \frac{v}{2\beta\Gamma\left(\frac{1}{\alpha}\right)} \exp\left(-\left(\frac{|x|}{\beta}\right)^v\right), \tag{4}$$

where $\beta = \alpha\sqrt{\Gamma\left(\frac{1}{v}\right)/\Gamma\left(\frac{3}{v}\right)}$ and $\Gamma(\cdot)$ is the gamma function with $\Gamma(a) = \int_0^\infty t^{a-1}e^{-t}dt$ and $a > 0$. In Eq. (4), the

shape parameter v is related to the 'shape' of the distribution, while σ^2 is related to the 'variance' of the distribution. We employ this GGD model to fit our MSCN coefficient distribution based on $\hat{I}$. These two GGD parameters (v, σ^2) are then regarded as the effective quality features.

As pointed out in [27], image quality information can also be captured by the distribution of the product of adjacent MSCN coefficient pairs. Inspired by the strategy from but in contrast to [27], as illustrated in Figure 3, we compute our MSCN coefficients along the major eight directions using

$M_{h1}(i,j) = \hat{I}(i,j)\hat{I}(i+1,j)$, $M_{d1}(i,j) = \hat{I}(i,j)\hat{I}(i+1,j-1)$, $M_{v1}(i,j) = \hat{I}(i,j)\hat{I}(i,j-1)$, $M_{d2}(i,j) = \hat{I}(i,j)\hat{I}(i-1,j-1)$, $M_{v2}(i,j) = \hat{I}(i,j)\hat{I}(i,j+1)$, $M_{d3}(i,j) = \hat{I}(i,j)\hat{I}(i+1,j+1)$, $M_{h2}(i,j) = \hat{I}(i,j)\hat{I}(i-1,j)$, $M_{d4}(i,j) = \hat{I}(i,j)\hat{I}(i-1,j+1)$.

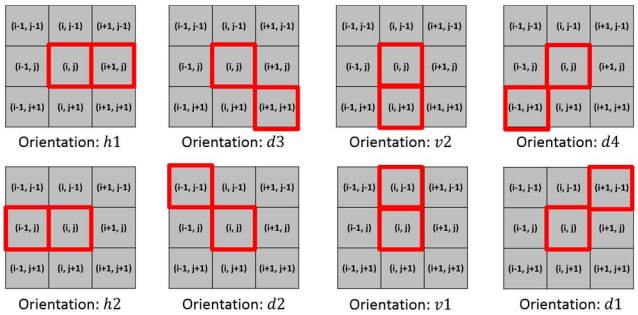

Figure 3: Illustration of the eight major directions for pair MSCN multiplication.

These eight paired product MSCN images are observed following a zero mode asymmetric generalized Gaussian distribution (AGGD) [28], which is defined as

$$f(x;\gamma,\alpha_l,\alpha_r) = \begin{cases} \dfrac{\gamma}{(\beta_l+\beta_r)\Gamma\left(\frac{1}{\gamma}\right)} \exp\left(-\left(\dfrac{-x}{\beta_l}\right)^{\gamma}\right), & \forall x \le 0 \\ \dfrac{\gamma}{(\beta_l+\beta_r)\Gamma\left(\frac{1}{\gamma}\right)} \exp\left(-\left(\dfrac{x}{\beta_r}\right)^{\gamma}\right), & \forall x > 0 \end{cases}, \quad (5)$$

where $\beta_l = \alpha_l\sqrt{\Gamma\left(\frac{1}{\gamma}\right)/\Gamma\left(\frac{3}{\gamma}\right)}$ and $\beta_r = \alpha_r\sqrt{\Gamma\left(\frac{1}{\gamma}\right)/\Gamma\left(\frac{3}{\gamma}\right)}$. In Eq. (5), the shape parameter γ adjusts the shape of the distribution while the scale parameters (α_l, α_r) adjust the spread of the left and right sides of the AGGD.

A promising feature is the mean parameter η that integrates three parameters using

$$\eta = (\beta_r - \beta_l)\frac{\Gamma(2/\gamma)}{\Gamma(1/\gamma)}. \quad (6)$$

Unlike the use of four parameters $(\gamma, \alpha_l, \alpha_r, \eta)$ of the AGGD in [27, 28], we only consider this mean parameter η as the quality feature. As multi-scale extraction of luminance enhances the accuracy of quality prediction, the features proposed in this section are computed twice but with the second time on a down-sampled by two and low-pass filtered image.

The objective of the proposed luminance aware features is to determine the luminance difference between real access faces and spoofing attack samples. These differences are based on the fact that the face skin has its own optical properties such as absorption, reflection, scattering, and refraction, while other materials such as paper, photographic paper, and electronic display exhibit different patterns. Accordingly, it is appropriate to employ the proposed features to characterize the structure and contrast information for face liveness detection.

3.2. Gradient magnitude similarity feature

The image gradient has been employed for IQA in different ways [29, 30]. The GMSD [6] is a recent IQA algorithm based on gradient magnitude similarity. Inspired by the GMSD, we develop an EPSD feature based on the improvement of the GMSD method. To increase the difference between the genuine and spoofing face images, the computation of our EPSD feature is restricted on apparent gradient positions, which are able to represent the characteristics of the whole gradient magnitude map.

The procedure starts from the computation of the horizontal and vertical gradients of the original subimage R as well as its distorted image D, which is obtained through the Gaussian filtering with a standard deviation σ_D. The gradient images are computed by convoluting the images with the Prewitt filter along two directions using

$$\mathrm{h}_x = \begin{bmatrix} \frac{1}{3} & 0 & \frac{-1}{3} \\ \frac{1}{3} & 0 & \frac{-1}{3} \\ \frac{1}{3} & 0 & \frac{-1}{3} \end{bmatrix}, \quad \mathrm{h}_y = \begin{bmatrix} \frac{1}{3} & \frac{1}{3} & \frac{1}{3} \\ 0 & 0 & 0 \\ \frac{-1}{3} & \frac{-1}{3} & \frac{-1}{3} \end{bmatrix}, \quad (7)$$

where h_x and h_y are the Prewitt filter kernels in the horizontal (x) and vertical (y) directions, respectively.

The gradient magnitude images of R and D at pixel locations $\mathrm{r}_i \in \mathrm{R}$ and $\mathrm{d}_i \in \mathrm{D}$ are computed using

$$\phi_r(\mathrm{r}_i) = \sqrt{(\mathrm{R}\otimes\mathrm{h}_x)^2(\mathrm{r}_i) + \left(\mathrm{R}\otimes\mathrm{h}_y\right)^2(\mathrm{r}_i)}, \quad (8)$$

$$\phi_d(\mathrm{d}_i) = \sqrt{(\mathrm{D}\otimes\mathrm{h}_x)^2(\mathrm{d}_i) + \left(\mathrm{D}\otimes\mathrm{h}_y\right)^2(\mathrm{d}_i)}, \quad (9)$$

where "$\otimes$" denotes the convolution operator, and $\phi_r(\mathrm{r}_i)$ and $\phi_d(\mathrm{d}_i)$ are the gradient magnitude images of R and D, respectively. The gradient magnitude similarity (GMS) map S is then calculated through a pixel-wise computation:

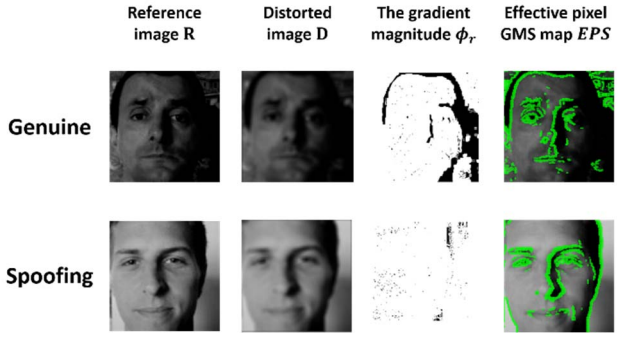

Figure 4: Illustration of feature differences between genuine and spoofing face images based on the EPS map.

$$S(i) = \frac{2\phi_r(i)\phi_d(i) + c}{\phi_r^2(i) + \phi_d^2(i) + c}, \tag{10}$$

where $i \in \phi_r, \phi_d$ and c is a constant that is used for numerical stability. The GMS map S is able to measure the distortion level at each pixel point of an image.

To more effectively separate spoofing images from genuine images, we consider pixels that have higher gradient magnitudes because low gradient magnitude pixels are likely to affect measurement and result in inaccuracy. Accordingly, a gradient magnitude criterion and the corresponding effective pixel similarity (EPS) map are defined as follows:

$$T_s = \min\{\text{first } p\% \text{ of sorted } S\}, \tag{11}$$

$$EPS(i) = \begin{cases} S(i) & \text{if } S(i) \geq T_s \\ 0 & \text{Otherwise} \end{cases}, \tag{12}$$

where p is a variable parameter corresponding to the number of pixels for selection based on the sorted GMS map S and T_s is regarded as a threshold for selecting the pixels that have higher gradient magnitude values to the EPS map as illustrated in Figure 4.

In order to generate a single quality score from the EPS map, a new deviation pooling method called *EPSD* is employed with

$$EPSD = \left(\frac{1}{N_S}\sum_{i=1}^{N_S}(EPS(i) - m_{EPS})^2\right)^{1/2}, \tag{13}$$

where m_{EPS} denotes the mean value of the EPS map and N_S is the number of pixels in the EPS map. Note that the value of the *EPSD* corresponds to the level of distortion in an image. The higher the *EPSD* score, the sharper the image. This implies that the proposed *EPSD* feature is suitable to detect the blurring effect generated during the re-quantization process of the original sample.

3.3. Feature descriptor

Images can be perceived as multi-scale and distortions affect image through different scales. Algorithms assessing the quality of an image perform better by incorporating with multi-scale information. The features proposed in Section 3.1 have included their downscaled counterparts. In addition to the *EPSD*, there are totally 21 perceptual features to construct a descriptor in this study. Let $\mathbf{z} = \{\upsilon_i\}_{i=1}^{21}$ be the descriptor vector containing all extracted features, which are listed in Table 1.

Table 1. Summary of the proposed perceptual features.

Feature ID	Feature Description	Computation Model
$\upsilon_1 - \upsilon_2$	(v, σ^2) (Eq. (4))	GGD
$\upsilon_3 - \upsilon_4$	(v, σ^2) (Eq. (4))	GGD (downscaled)
$\upsilon_5 - \upsilon_{12}$	η (Eq. (6))	AGGD
$\upsilon_{13} - \upsilon_{20}$	η (Eq. (6))	AGGD (downscaled)
υ_{21}	*EPSD* (Eq. (13))	EPSD

3.4. Classification

The LibSVM library of the support vector machine [31] is utilized for classification. After concatenating the image quality features to construct the feature descriptor based on the training data, we can create a trained SVM model through the feature descriptor. By using the trained SVM model, we are able to detect liveness spoofing face images based on the testing data and make a decision.

4. Experimental Results and Analysis

In this section, we present the experiments performed and the results obtained to demonstrate the effectiveness of the proposed algorithm. Section 4.1 details the datasets adopted for our experiments and Section 4.2 introduces the experimental protocols followed in this paper. Section 4.3 describes the implementation details of the proposed method and Section 4.4 compares the performance of our algorithm with state of the art methods.

4.1. Benchmark datasets

To assess the effectiveness of our proposed method, we consider three latest datasets:

- *CASIA Face Anti-Spoofing Dataset [8]:* This dataset contains video recordings of genuine and spoofing faces. The real faces were recorded based on 50 genuine subjects, whereas the spoofing faces were made from high-quality recordings of the genuine faces. The data were divided into three categories according to the video quality: low, normal, and high. The low-quality and normal-quality videos were captured by a USB camera with 480x640 pixels, whereas the high-quality videos were captured with a Sony NEX-5 camera with 1920x1080 pixels.
- *Replay-Attack Dataset [7]:* This dataset consists of 1200 short video recordings of both real-access and attack attempts of 50 different individuals. There are

200 valid access videos, 200 printed photo-based video attacks, 400 mobile-based video attacks, and 400 high-resolution video attacks.

- *UVAD Dataset [9]:* This dataset comprises 808 valid access videos and 16268 attempted attack videos of 404 different people. Seven different devices were utilized to capture the attempted videos with full HD quality.

4.2. Experimental protocol

To evaluate the discrimination rate, we employed the following measures: false acceptance rate (FAR) and false rejection rate (FRR). The FAR is a measure for the number of face images misclassified as real, whereas the FRR is a measure for the number of real images misclassified as fake. In addition to the equal error rate (EER) and the relative error reduction (RER), the half total error rate (HTER) is computed using HTER=((FAR+FRR))⁄2, which is a more objective measure to evaluate the proposed method. We follow the experimental protocols adopted in the Replay-Attack, CASIA, and UVAD datasets to design the experiments.

(i) *Protocol 1 with CASIA*: the data are divided into two subsets: training set and testing set. Based on the recommendation in [8], the training set is utilized to train a classifier and the EER values are subsequently generated using the testing set.

(ii) *Protocol 2 with Replay-Attack*: the data are separated into three subsets: training set, development set, and testing set [7]. The training set is employed to build the model, the development set for tuning the parameters, and the testing set to compute the final HTER.

(iii) *Protocol 3 with UVAD:* the UVAD data contain six subsets with both valid access and attempted attack videos, depending on different acquisition sensor devices and brands [12]. We train the model using videos captured with devices of Sony, Olympus, and Kodak, and test the model using videos with devices of Canon, Nikon, and Panasonic.

4.3. Implementation details

In our proposed algorithm, there are several parameters that need to be decided. To determine the parameter values more effectively, we tuned the parameters based on the training subset of the Replay-Attack dataset [7]. The parameter value leading to the lowest HTER value is chosen. For the luminance aware feature, the values of μ_I and σ_I were obtained by using a 7 x 7 local window with a 2D circularly-symmetric Gaussian weighting function sampled out to three standard deviation ($K = L = 3$) and rescaled to unit volume. The parameters related to the EPSD are set as follows: $p = 5$ and $\sigma_D = 0.6$, where p corresponds to the magnitude threshold in Eq. (11) and σ_D is a critical parameter in the Gaussian filter for the distorted image D.

4.4. Performance analysis

We compared the proposed method with various state of the art methods based on the Replay-Attack, CASIA, and UVAD datasets. As presented in Table 2, our algorithm outperformed other methods with the lowest score of HTER=5.38% on the Replay-Attack dataset. The evaluation was based on texture analyses, motion analyses, and fusion schemes. Table 3 presents our face liveness detection results in comparison with the state of the art methods based on the CASIA dataset. It was noted that the proposed method achieved the second best result with EER=12.7% comparing to the best EER of 10.0%. As demonstrated in Table 4, our proposed algorithm produced the smallest scores of HTER=23.00% and FRR=9.20% on the UVAD dataset. While the LBP [7] provided the lowest score of FAR=27.41%, its other evaluation metrics were pretty high with HTER=46.72% and FRR=66.04%.

Table 2. Performance comparison between different methods in terms of the HTER and RER based on the Replay-Attack dataset protocol.

Methods	HTER (%)	RER (%)
Chingovska et al. [7]	15.16	64.51
Allan Pinto et al. [12]	14.27	62.30
Maatta et al. [14]	13.87	61.21
Anjos and Marcel et al. [32]	11.79	54.37
Gragnaniello et al. [2]	9.40	42.77
Wen et al. [21]	7.40	27.30
Pereira et al. [33]	7.60	29.21
Proposed	**5.38**	**0**

Table 3. Performance comparison between different methods in terms of the EER based on the CASIA dataset protocol.

Methods	EER (%)
DoG [8]	17.0
LBP [33]	16.0
LBP-TOP [33]	**10.0**
Motion mag+LBP [34]	14.4
Pinto et al. [12]	14.0
DMD [16]	21.8
Proposed	12.7

Table 4. Performance comparison between different methods in terms of the FAR, FRR, and HTER based on the UVAD dataset.

Methods	FAR (%)	FRR (%)	HTER (%)
LBP [7]	**27.41**	66.04	46.72
Correlation [32]	81.60	14.56	48.06
Pinto et al. [12]	44.73	15.00	29.87
Proposed	36.80	**9.20**	**23.00**

Table 5. Performance evaluation of the proposed method based on the cross-dataset protocols.

Train	Test	FAR (%)	FRR (%)	HTER (%)	Mean HTER (%)
Replay-Attack	CASIA	75.83	2.22	39.03	43.82
	UVAD	75.60	21.60	48.60	
CASIA	Replay-Attack	0.00	76.25	38.13	50.97
	UVAD	36.00	91.60	63.80	
UVAD	CASIA	97.50	8.89	53.19	53.22
	Replay-Attack	49.50	57.00	53.25	

Table 6. Performance comparison between different methods in terms of the HTER based on the cross-dataset protocols.

Methods	Train	Test	HTER (%)
Motion-Mag	Replay-Attack	CASIA	47.0
	CASIA	Replay-Attack	50.1
Pinto et al.	Replay-Attack	CASIA	50.0
	CASIA	Replay-Attack	**34.4**
LBP	Replay-Attack	CASIA	57.1
	CASIA	Replay-Attack	47.1
LBP-TOP	Replay-Attack	CASIA	61.3
	CASIA	Replay-Attack	50.6
Correlation	Replay-Attack	CASIA	48.3
	CASIA	Replay-Attack	50.3
Proposed	Replay-Attack	CASIA	**39.0**
	CASIA	Replay-Attack	38.1

Table 5 summarizes the cross-dataset evaluation results of the proposed method based on the Replay-Attack, CASIA, and UVAD datasets. It was noted that the experiments with the Replay-Attack dataset for training produced the lowest HTER score of 43.82%.This is probably because that the Replay-Attack dataset has more diverse dataset types such as printed, phone, and tablet attack videos. Thanks to the designed luminance and gradient features in our framework, these variations in the Replay-Attack dataset were appropriately resolved in the training phase, which in turn generated robust detection abilities as shown. Finally, in Table 6, we compared our cross-dataset results with the state of the art methods. It was observed that the proposed algorithm achieved the smallest HTER=39.0% on the experiments of the Replay-Attack for training and the CASIA for testing. In the scenario of the reverse experiments, Pinto et al. [12] methods produced the lowest HTER=34.4%. However, they had a much higher HTER score of 50.0% for the other protocol. Moreover, our method exhibited quite close and low HTER scores on both protocols, which were not observed in other methods.

5. Conclusion

In this study, we have proposed an effective approach against video-based face spoofing attacks based on perceptual image quality assessment features with multi-scale analysis. The proposed framework took advantage of the BIQE, EPSD, and GMS by selecting effective pixels to create the most appropriate image quality features for face liveness detection. We have conducted extensive experiments to evaluate the performance of the proposed method and have compared against various state of the art methods. The experimental results indicated that our proposed framework achieved relatively smaller scores in terms of the adopted error evaluation metrics on the Replay-Attack, CASIA, and UVAD datasets, which outperformed many state of the art methods. We believe that the incorporation of the image quality assessment knowledge into face liveness detection is promising to improve the overall accuracy. Nevertheless, there are some aspects that can be further investigated. For example, we only considered the case of video-based spoofing attacks. In the future, 3D mask attacks are worthy taking into account.

Acknowledgements

This work was supported by the Ministry of Science and Technology of Taiwan under Grant No. MOST 106-2221-E-002-082.

References

[1] A. A. Ross, K. Nandakumar, and A. Jain, *Handbook of multibiometrics*. Springer Science & Business Media, 2006.

[2] D. Gragnaniello, G. Poggi, C. Sansone, and L. Verdoliva, "An investigation of local descriptors for biometric spoofing detection," *IEEE transactions on information forensics and security,* vol. 10, no. 4, pp. 849-863, 2015.

[3] S. A. Schuckers, "Spoofing and anti-spoofing measures," *Information Security technical report,* vol. 7, no. 4, pp. 56-62, 2002.

[4] K. Delac and M. Grgic, "A survey of biometric recognition methods," in *Electronics in Marine, 2004. Proceedings Elmar 2004. 46th International Symposium*, 2004, pp. 184-193: IEEE.

[5] M. Pantic, A. Pentland, A. Nijholt, and T. S. Huang, "Human computing and machine understanding of human behavior: A survey," in *Artifical Intelligence for Human Computing*: Springer, 2007, pp. 47-71.

[6] W. Xue, L. Zhang, X. Mou, and A. C. Bovik, "Gradient magnitude similarity deviation: A highly efficient perceptual image quality index," *IEEE Transactions on Image Processing,* vol. 23, no. 2, pp. 684-695, 2014.

[7] I. Chingovska, A. Anjos, and S. Marcel, "On the effectiveness of local binary patterns in face anti-spoofing," in *Biometrics Special Interest Group (BIOSIG), 2012 BIOSIG-Proceedings of the International Conference of the*, 2012, pp. 1-7: IEEE.

[8] Z. Zhang, J. Yan, S. Liu, Z. Lei, D. Yi, and S. Z. Li, "A face antispoofing database with diverse attacks," in *Biometrics (ICB), 2012 5th IAPR international conference on*, 2012, pp. 26-31: IEEE.

[9] A. Pinto, W. R. Schwartz, H. Pedrini, and A. de Rezende Rocha, "Using visual rhythms for detecting video-based facial spoof attacks," *IEEE Transactions on Information Forensics and Security,* vol. 10, no. 5, pp. 1025-1038, 2015.

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

[11] S. Marcel, M. S. Nixon, and S. Z. Li, *Handbook of Biometric Anti-Spoofing*. Springer, 2014.

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

[13] H. Steiner, A. Kolb, and N. Jung, "Reliable face anti-spoofing using multispectral SWIR imaging," in *Biometrics (ICB), 2016 International Conference on*, 2016, pp. 1-8: IEEE.

[14] J. Määttä, A. Hadid, and M. Pietikäinen, "Face spoofing detection from single images using micro-texture analysis," in *Biometrics (IJCB), 2011 international joint conference on*, 2011, pp. 1-7: IEEE.

[15] X. Tan, Y. Li, J. Liu, and L. Jiang, "Face liveness detection from a single image with sparse low rank bilinear discriminative model," *Computer Vision–ECCV 2010,* pp. 504-517, 2010.

[16] S. Tirunagari, N. Poh, D. Windridge, A. Iorliam, N. Suki, and A. T. Ho, "Detection of face spoofing using visual dynamics," *IEEE transactions on information forensics and security,* vol. 10, no. 4, pp. 762-777, 2015.

[17] S. R. Arashloo, J. Kittler, and W. Christmas, "Face spoofing detection based on multiple descriptor fusion using multiscale dynamic binarized statistical image features," *IEEE Transactions on Information Forensics and Security,* vol. 10, no. 11, pp. 2396-2407, 2015.

[18] Z. Boulkenafet, J. Komulainen, and A. Hadid, "Face spoofing detection using colour texture analysis," *IEEE Transactions on Information Forensics and Security,* vol. 11, no. 8, pp. 1818-1830, 2016.

[19] P. P. Chan *et al.*, "Face Liveness Detection Using a Flash against 2D Spoofing Attack," *IEEE Transactions on Information Forensics and Security,* 2017.

[20] J. Yang, Z. Lei, D. Yi, and S. Z. Li, "Person-specific face antispoofing with subject domain adaptation," *IEEE Transactions on Information Forensics and Security,* vol. 10, no. 4, pp. 797-809, 2015.

[21] D. Wen, H. Han, and A. K. Jain, "Face spoof detection with image distortion analysis," *IEEE Transactions on Information Forensics and Security,* vol. 10, no. 4, pp. 746-761, 2015.

[22] K. Patel, H. Han, and A. K. Jain, "Secure face unlock: Spoof detection on smartphones," *IEEE Transactions on Information Forensics and Security,* vol. 11, no. 10, pp. 2268-2283, 2016.

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

[24] D. Menotti *et al.*, "Deep representations for iris, face, and fingerprint spoofing detection," *IEEE Transactions on Information Forensics and Security,* vol. 10, no. 4, pp. 864-879, 2015.

[25] Y. Atoum, Y. Liu, A. Jourabloo, and X. Liu, "Face Anti-Spoofing Using Patch and Depth-Based CNNs."

[26] D. L. Ruderman and W. Bialek, "Statistics of natural images: Scaling in the woods," in *Advances in neural information processing systems*, 1994, pp. 551-558.

[27] A. Mittal, A. K. Moorthy, and A. C. Bovik, "No-reference image quality assessment in the spatial domain," *IEEE Transactions on Image Processing,* vol. 21, no. 12, pp. 4695-4708, 2012.

[28] N.-E. Lasmar, Y. Stitou, and Y. Berthoumieu, "Multiscale skewed heavy tailed model for texture analysis," in *Image Processing (ICIP), 2009 16th IEEE International Conference on*, 2009, pp. 2281-2284: IEEE.

[29] D.-O. Kim, H.-S. Han, and R.-H. Park, "Gradient information-based image quality metric," *IEEE Transactions on Consumer Electronics,* vol. 56, no. 2, 2010.

[30] A. Liu, W. Lin, and M. Narwaria, "Image quality assessment based on gradient similarity," *IEEE Transactions on Image Processing,* vol. 21, no. 4, pp. 1500-1512, 2012.

[31] C.-C. Chang and C.-J. Lin, "LIBSVM: a library for support vector machines," *ACM transactions on intelligent systems and technology (TIST),* vol. 2, no. 3, p. 27, 2011.

[32] A. Anjos and S. Marcel, "Counter-measures to photo attacks in face recognition: a public database and a baseline," in *Biometrics (IJCB), 2011 international joint conference on*, 2011, pp. 1-7: IEEE.

[33] T. de Freitas Pereira *et al.*, "Face liveness detection using dynamic texture," *EURASIP Journal on Image and Video Processing,* vol. 2014, no. 1, p. 2, 2014.

[34] S. Bharadwaj, T. I. Dhamecha, M. Vatsa, and R. Singh, "Computationally efficient face spoofing detection with motion magnification," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops*, 2013, pp. 105-110.