Hindawi Publishing Corporation
Journal of Electrical and Computer Engineering
Volume 2016, Article ID 4721849, 14 pages
http://dx.doi.org/10.1155/2016/4721849

Research Article

Face Spoof Attack Recognition Using Discriminative Image Patches

Zahid Akhtar and Gian Luca Foresti

Department of Mathematics and Computer Science, University of Udine, Via delle Scienze 206, 33100 Udine, Italy

Correspondence should be addressed to Zahid Akhtar; zahid.akhtar@uniud.it

Received 3 December 2015; Revised 12 March 2016; Accepted 14 April 2016

Academic Editor: Kwok-Wai Cheung

Face recognition systems are now being used in many applications such as border crossings, banks, and mobile payments. The wide scale deployment of facial recognition systems has attracted intensive attention to the reliability of face biometrics against spoof attacks, where a photo, a video, or a 3D mask of a genuine user's face can be used to gain illegitimate access to facilities or services. Though several face antispoofing or liveness detection methods (which determine at the time of capture whether a face is live or spoof) have been proposed, the issue is still unsolved due to difficulty in finding discriminative and computationally inexpensive features and methods for spoof attacks. In addition, existing techniques use whole face image or complete video for liveness detection. However, often certain face regions (video frames) are redundant or correspond to the clutter in the image (video), thus leading generally to low performances. Therefore, we propose seven novel methods to find discriminative image patches, which we define as regions that are salient, instrumental, and class-specific. Four well-known classifiers, namely, support vector machine (SVM), Naive-Bayes, Quadratic Discriminant Analysis (QDA), and Ensemble, are then used to distinguish between genuine and spoof faces using a voting based scheme. Experimental analysis on two publicly available databases (Idiap REPLAY-ATTACK and CASIA-FASD) shows promising results compared to existing works.

1. Introduction

In the last years, face recognition systems have gained interest due to face's rich features that offer a strong biometric cue to recognize individuals for a wide variety of applications in both law and nonlaw enforcements [1]. In fact, facial recognition systems are already in operation worldwide, including USVISIT, which is a US Customs and Border Protection (CBP) management system, UIDAI that provides identity to all persons resident in India, and Microsoft Kinect which uses face recognition to access dashboard and automatic sign-in to Xbox Live profile. Similarly, face biometrics is also nowadays being used ubiquitously as an alternative to passwords on mobile devices such as Android KitKat mobile OS, Lenovo VeriFace, Asus SmartLogon, and Toshiba SmartFace.

Despite the great deal of progress in facial recognition systems, vulnerabilities to face spoof attacks are mainly overlooked [2]. Facial spoof attack is a process in which a fraudulent user can subvert or attack a face recognition system by masquerading as registered user and thereby gaining illegitimate access and advantages [1, 3–5]. Face spoofing attack is a major issue for companies selling face biometric-based identity management solutions [6]. For instance, at New York City, nonwhite robbers disguised themselves as white cops, using life-like latex masks, and were caught robbing a cash-checking store in 2014 (see Figure 1, also for other recent face spoof attacks).

Recent study reported in [1] suggests that the success rate of face spoof attacks could be up to 70%, even when a state-of-the-art Commercial Off-The-Shelf (COTS) face recognition system is used. Therefore, we could infer that even COTS face recognition systems are mainly not devised to effectively distinguish spoof faces from genuine live faces. As a matter of fact, this vulnerability of face spoofing to face recognition systems is now enlisted in the National Vulnerability Database of the National Institute of Standards and Technology (NIST) in the US.

Typical countermeasure to face spoof attacks is liveness detection method, which aims at disambiguating human live face samples from spoof artifacts [2, 7]. There exist

(a) (b)

Figure 1: (a) In New York, a robber disguised as white cop using latex masks robbing a cash-checking store. (b) In 2010, a passenger boarded a plane in Hong Kong with an old man mask and arrived in Canada to claim asylum.

several face antispoofing or liveness detection techniques [7–15]. However, face spoofing attacks remain a problem due to difficulties in finding discriminative and computationally inexpensive features and techniques for spoof recognition. Moreover, published methods are limited in their scope since they mainly use whole face image or complete video for liveness detection. Nevertheless, often certain face image regions (video frames) are redundant or correspond to the clutter in the image (video), leading thus generally to low performances.

It is thus essential to develop robust, efficient, and compact face antispoofing (or liveness detection) methods, which are capable of generalizing well to discriminative, class-specific information and imaging conditions. To this aim, in this paper, we propose a simple and effective solution based on discriminative image patches. In particular, we propose seven novel fully automated algorithms to highlight regions of interest in face images. We define these regions (or image patches) to be discriminative (i.e., specific to a particular class: live or spoof), consistent (i.e., reliably appearing in different face images or video frames), salient (i.e., conspicuous regions), and repetitive (i.e., frequently appearing in the image set of specific class). The basic notion is "interesting patches are those that are specific to a face image (or video frame) and should contain features that give assistance to discriminate a given live face image from spoofed one." Based on this definition, two of the seven proposed image patch selection methods (i.e., MAXDIST and DEND-CLUSTER) do not employ any training or prior learning. However, the remaining techniques use simple clustering (i.e., CP and CS), image intensity (i.e., IPI), image quality (i.e., IQA), or diversity filter (i.e., DF) to obtain discriminative patches. For final classification, we exploited four well-known classifiers, namely, support vector machine (SVM), Naive-Bayes, Quadratic Discriminant Analysis (QDA), and Ensemble, using voting based scheme. Experimental analysis on two publicly available databases (Idiap Replay-Attack and CASIA-FASD) shows good results compared to existing works.

The added advantages of the proposed framework are (i) being cheap, (ii) very low complexity, (iii) needing one face image (i.e., the same face sample acquired for face recognition) to detect whether it is genuine or spoof attack, (iv) being nonintrusive, (v) being user-friendly, and (vi) being easy to embed in already functional face recognition systems plus no requirement of new piece of hardware.

The reminder of the paper is organized as follows. Exiting literature works on face liveness detection are discussed in Section 2. The proposed approaches to determine the discriminative image patches and spoof classification schemes are described in Section 3. Experimental datasets, protocols, and results are presented in Section 4. A conclusion is drawn in Section 5.

2. Related Work

Despite great deal of advancements in face recognition systems, face spoofing still poses a serious threat. Most of the existing academic and commercial facial recognition systems may be spoofed by (see Figure 2) (i) a photo of a genuine user; (ii) a video of a genuine user; (iii) a 3D face model (mask) of a genuine user; (iv) a reverse-engineered face image from the template of a genuine user; (v) a sketch of a genuine user; (vi) an impostor wearing specific make-up to look like a genuine user; (vii) an impostor who underwent plastic surgery to look like a genuine user. The most easiest, cheapest, and common face spoofing attack is to submit a photograph of a legitimate user to the face recognition systems, which is also known as "photo attack."

Typical countermeasure (i.e., face liveness detection or antispoofing) techniques can be coarsely classified in three categories based on clues used for spoof attack detection: (i) *motion analysis based methods*, (ii) *texture analysis based methods*, and (iii) *hardware-based methods*. In what follows, we provide a brief literature overview of published face spoof recognition techniques along with their pros and cons.

(i) Motion Analysis Based Methods. These methods broadly try to detect spontaneous movement clues generated when two dimensional counterfeits are presented to the camera of the system, such as photographs or videos. Therefore, Pan et al. [7] exploited the fact that human eye-blink occurs once every 2–4 seconds and proposed eye-blink based liveness detection for photo-spoofing using (spontaneous) eye-blinks. This method uses an undirected conditional random field framework to model the eye-blinking, which relaxes the independence assumption of generative modelling and states dependence limitations from hidden Markov modelling. It is evident that real human faces (which are 3D objects) will move significantly differently from planer objects, and such deformation patterns can be employed for liveness detection. For example, Tan et al. [8] considered Lambertian

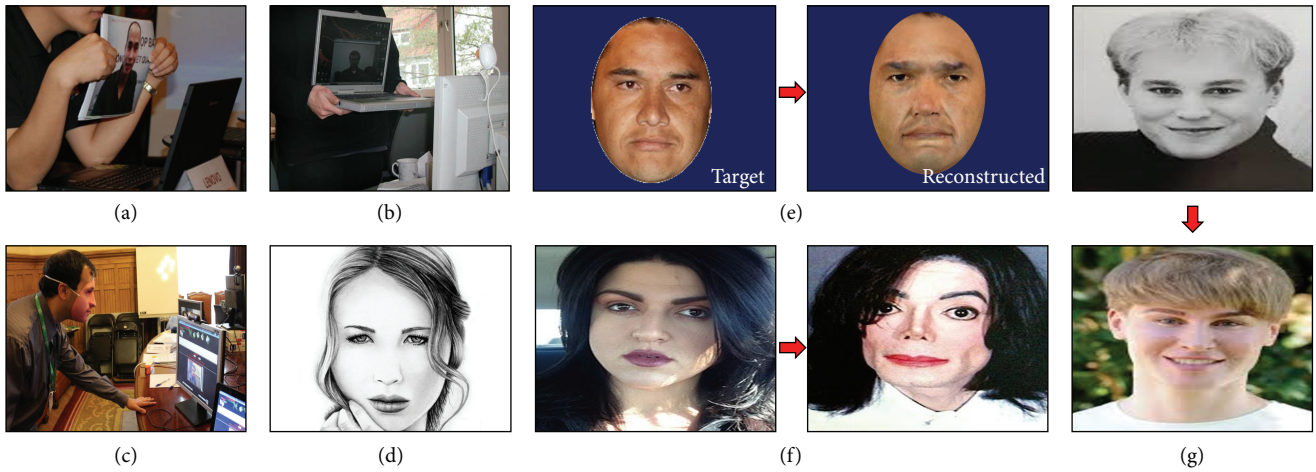

Figure 2: Examples of face spoofing using (a) photograph, (b) video, (c) 3D mask, (d) sketch, (e) reverse-engineered face image, (f) make-up (skillful application of make-up to look like Michel Jackson), and (g) plastic surgery (this boy underwent excessive plastic surgery to look like Justin Bieber).

reflectance model with difference-of-Gaussians (DoG) to derive differences of motion deformation patterns between 2D face photos presented during spoofing attacks and 3D live faces. It does so by extracting the features in both scenarios using a variational retinex-based method and difference-of-Gaussians (DoG) [9] based approach. Then, the features are used for live or spoof classification. Reported experiments showed promising results on a dataset consisting of real accesses and spoofing attacks to 15 clients using photo-quality and laser-quality prints. Kollreider et al. [10] proposed a liveness detection approach based on a short sequence of images using a binary detector, which captures and tracts the subtle movements of different selected facial parts using a simplified optical flow analysis followed by a heuristic classifier. The same authors also presented a method to combine scores from different experts systems, which concurrently observe the 3D face motion approach introduced in the former work as liveness attributes like eye-blinks or mouth movements. In the similar fashion, Bao et al. [11] also used optical flow to estimate motion for detecting attacks produced with planar media such as prints or screens.

Since the frequency of facial motion is restricted by the human physiological rhythm, thus motion based methods take a relatively long time (usually > 3 s) to accumulate stable vitality features for face spoof detection. Moreover, they may be circumvented or confused by other motions, for example, background motion in the video attacks.

(ii) Texture Analysis Based Methods. This kind of methods examines the skin properties, such as skin texture and skin reflectance, under the assumption that surface properties of real faces and prints, for example, pigments, are different. Examples of detectable texture patterns due to artifacts are printing failures or blurring. Li et al. [13] described a method for print-attack face spoofing by exploiting differences in the 2D Fourier spectra of live and spoof images. The method assumes that photographs are normally smaller in size and contain fewer high frequency components compared to real faces. The method only works well for downsampled photos of the attacked identity but likely fails for higher-quality samples. In [14, 16, 17], authors developed microtexture analysis based methods to detect printed photo attacks. One limitation of presented methods is the requirement of reasonably sharp input image. Recently, Galbally et al. [3] designed a face spoof detection scheme based on 25 different image quality measures: 21 full-reference measures and 4 nonreference measures. However, all 25 image quality measures are required to get good results and no face-specific information has been considered in designing informative features for face spoof detection.

Compared to other techniques, texture analysis based algorithms are generally faster to classify a spoof attack. Nevertheless, they could be easily overfitted to one particular illumination and imagery condition and hence do not generalize well to different spoofing conditions.

(iii) Hardware-Based Methods. Few interesting hardware-based face antispoofing techniques have been proposed so far based on imaging technology outside the visual spectrum, such as 3D depth [18], complementary infrared (CIR), or near infrared (NIR) images [15] by comparing the reflectance information of real faces and spoof materials using a specific set-up of LEDs and photodiodes at two different wavelengths. Preliminary efforts on thermal imaging for face liveness detection have also been exploited, including the acquisition of large database of thermal face images for real and spoofed access attempts [19]. Besides, numbers of researchers have explored multimodality as antispoofing techniques for face spoofing attacks. They have mainly considered the combination of face and voice by utilizing the correlation between the lips movement and the speech being produced [20], where a microphone and a speech analyzer were required. Similarly, challenge-response strategy considering voluntary eye-blinking and mouth movement following a request from the system has been studied in [10]. Though hardware-based methods provide better results and performances, they

Table 1: Summary of different face spoof detection techniques.

Method	Main features used	Pros and cons
Motion analysis based methods	Motion detection [10] Eye-blink detection using conditional random fields (CRF) [7] Face motion detection using Optical Flow Lines (OFL) [10] Context-based using correlation between face motion and background motion [17]	(i) Good generalization capability (ii) High computational cost (iii) Easily circumvented by fake motions
Texture analysis based methods	Face texture using Lambertian model [8] Face texture using LBPs [17] Texture + shape combining LBPs + Gabor Wavelets + HOG [5]	(i) Fast response (ii) Low computational cost (iii) Poor generalization capability
Hardware-based methods	Multimodality: face and voice [20] Thermal images [19] Reflectance in 3D [21]	(i) Better generalization capability (ii) Extra hardware requirement (iii) High cost of the system

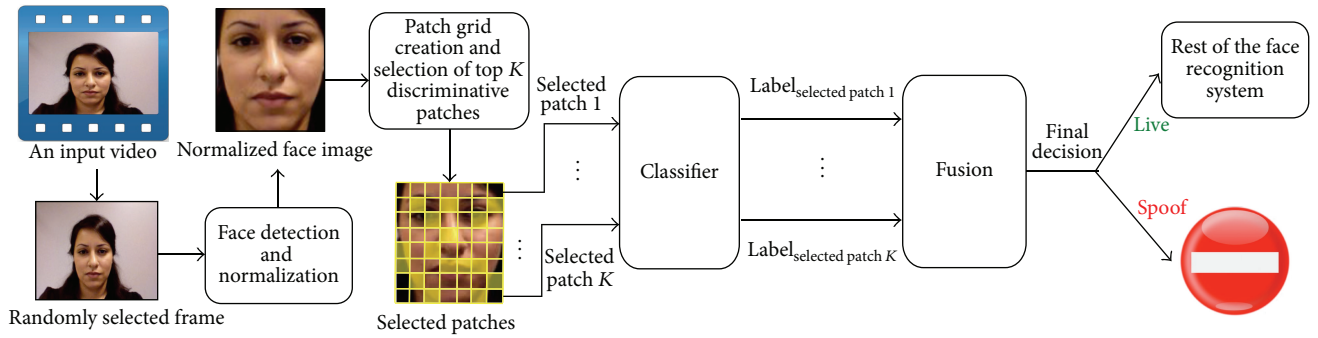

Figure 3: The proposed face spoof recognition algorithm based on discriminative image patches.

require extra piece of hardware which increases the cost of the system. A summary with relevant features of the most representative works in face antispoofing is presented in Table 1.

Though there exist several face antispoofing or liveness detection techniques, face spoof attacks remain an issue because of difficulty in finding discriminative and computationally inexpensive features and mechanisms for spoof recognition. Reported methods are limited in their scope since they mainly use full image or complete video for liveness detection. In particular, there is a lack of investigation on how specific image patches rather than full image perform in spoof detection. As often image regions (video frames) are redundant or correspond to the clutter in the image (video), leading thus generally to low performances and high computational cost, towards this direction, we propose seven novel methods to find discriminative image patches, which we define as regions that are salient, instrumental, and class-specific. Experimental results show that the proposed methods obtain comparable performances to existing techniques.

3. The Proposed Method for Face Spoof Recognition

Face spoof detection can be seen as a two-class classification problem, where the input face image (or video) has to be flagged as either live or spoof. The keynote of the process is attaining a discriminant feature set together with an appropriate classification scheme that gives the probability of the image (or video) realism. Practical face spoof detection requires that a decision be made based on single image or a limited number of frames in the video-based system. In this work, thus we aim to design simple but effective solution based on discriminative image patches using a single face frame/image. We define these image patches to be discriminative, consistent, salient, and repetitive. The notion is that the interesting patches are those that are specific to a face image (or video frame) and should contain features that help discriminate a given live face image from spoofed one.

Figure 3 shows the schematic diagram of the proposed face spoof recognition algorithm based on discriminative image patches. The proposed framework first randomly selects a single frame from a given face video (in case of image-based system, the given single image is used). Then, face is detected using Local SMQT Features and Split-Up Snow Classifier [30]. Subsequently, the detected face is densely divided into a grid of nonoverlapping local patches. These patches are ranked based on their discriminative power. The top K patches are selected using specific discriminative image patch selection method among the proposed techniques (explained below in detail). For each selected image patch, features are extracted that are then fed into particular classifier (i.e., SVM, Naive-Bayes, QDA, or

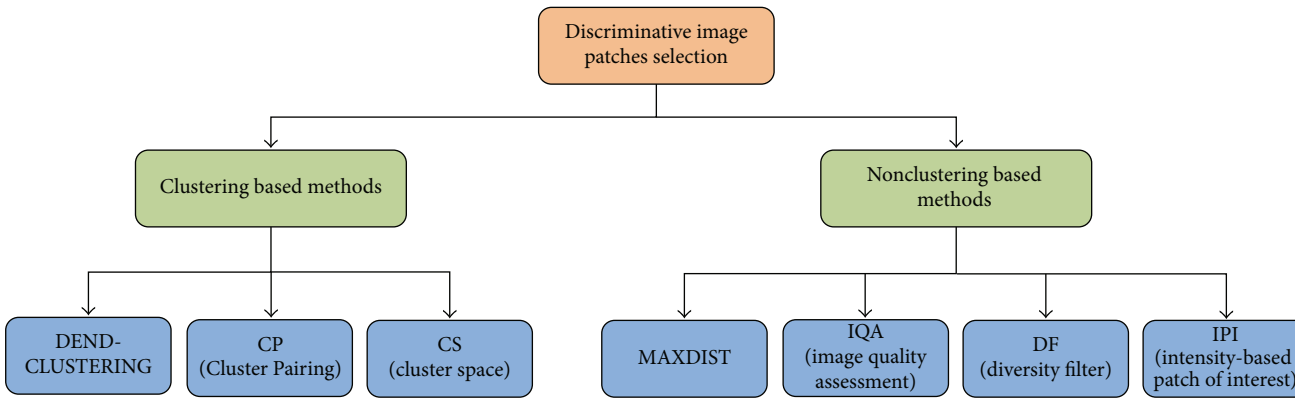

Figure 4: Classification of the seven discriminative image patches selection methods proposed in this work.

Ensemble classifier). The classification results of individual patches are combined by a majority-voting based scheme to obtain the final binary decision: genuine or spoof face.

3.1. Discriminative Image Patches Selection Methods. In what follows, we give the details of the proposed seven methods for discriminative image patches selection. The proposed patch selection methods are grouped into two categories: clustering based methods and nonclustering based methods. For clarity, in Figure 4 we show a diagram with the patch selection methods classification followed in this section.

3.1.1. Clustering Based Methods. The patch selection methods in this category rely on a clustering algorithm at any specific stage of the procedure. Three clustering based methods proposed in this work are as follows.

(1) DEND-CLUSTERING. In this discriminative image patches selection technique, the n patches in the given face image are grouped into $\mathbf{T}$ clusters, such that patches within a cluster are more similar to each other than patches belonging to different clusters. Then, for each cluster a prototype (representative) patch that typifies the members of that cluster is chosen, resulting in $\mathbf{T}$ discriminative patches. Since this method uses the dendrogram [31] to choose the discriminative patches, thus we call it DEND-CLUSTERING.

For each normalized face image ($\mathbf{I}$), first a set of dense patches $\{\mathbf{P}_i(\mathbf{I}) \in \mathbb{R}^{M\times N}\}_{i=1}^{n}$ is taken, where n is the total number of dense patches. Computation of the dissimilarity scores between patches is needed to perform clustering. Therefore, first the dissimilarity between successive patches is computed by comparing the respective features. Then, hierarchical clustering [31] is exploited, because the representation of the n patches is in the form of an $n \times n$ dissimilarity matrix instead of an $n \times p$ pattern matrix, where p is the dimension of feature vector. In particular, an agglomerative complete link clustering algorithm [31] is used in this work. The outcome of this algorithm is a dendrogram (a binary tree), where each terminal node corresponds to a patch, and the intermediate nodes indicate the formation of clusters. The discriminative K patches are selected as follows:

(i) Find the pairwise distance scores between the n patches to form the dissimilarity matrix $\mathbf{D}$.

(ii) Apply the complete link clustering algorithm on $\mathbf{D}$, and generate the dendrogram $\mathbf{L}$. Use the dendrogram $\mathbf{L}$ to identify $\mathbf{T}$ clusters. The method in [31] automatically determines the threshold distance to cut the dendrogram and identify exactly $\mathbf{T}$ clusters.

(iii) In each of the clusters identified in step (ii), select a patch whose average distance from the rest of the patches in the cluster is minimum. If a cluster has only 2 patches, choose any one of the two patches at random.

(iv) The patches selected in step (iii) are arranged in descending order on the basis of their ideal selection measure (ISM) value, which is computed as

$$\mathrm{ISM}(\mathbf{P}) = \sum_{x=1}^{P-2}\sum_{y=1}^{Q-2} G(x,y), \tag{1}$$

where $\mathbf{P}$ is a patch of size $P \times Q$ and $G(x, y)$ is the image gradient at location (x, y).

(v) The top K patches are selected as discriminative patches.

It is worth mentioning that steps (i)–(iii) in DEND-CLUSTERING method have close resemblance with the technique in [31] for fingerprint template selection. Here, we extended the technique by proposing step (iv) to be utilized for ranking and selection of discriminative patches.

(2) CP (Cluster Pairing). Apparently, the discrimination power of patches (features) decides maximum possible classification accuracy, and thus prior knowledge of "how cluttered the features (patches) may be" and "their contribution to classes separability in the feature space" can help to design and accomplish better classification scheme and accuracy. To this aim, in this method, first two independent sets of clusters are generated using genuine and spoof attack samples, respectively. Since overlapping of interclass clusters has great effect on classification accuracy, therefore pairs

of overlapped corresponding clusters of two independent sets are formed using minimum distance between them. Finally, patches which do not belong to both clusters of a given interclass highly overlapped cluster pair are selected as discriminative patches. In other words, if a patch belongs to both clusters of a given interclass cluster pair, it means that its features cause overlapping different classes in the feature space, which might thus lead to lower classification accuracy. The steps followed to obtain top K discriminative patches using this method are as follows:

(i) Two independent sets of clusters are generated using live and spoof attack training samples, respectively. Each class is represented by X number of clusters scattered in the feature space. K-means clustering technique is exploited in this work for cluster creation.

(ii) All possible pairs of corresponding clusters of two independent sets are formed using

$$D_{\text{mean}}\left(C_i, C_j\right) = \left\|m_i - m_j\right\|_{j=1}^{X} \leq \eta, \tag{2}$$

where η is threshold, m_x is center of C_x, C_i is a given cluster from live class clusters set, and C_j is a given cluster from spoof class clusters set. D_{mean} is appropriate for detecting spherical and compact cluster pairs, since each cluster is represented only by its center point.

(iii) For a given face image, K patches are chosen as discriminative patches, which do not belong to both elements of the interclass clusters.

(3) CS (Cluster Space). Principally, information classes cannot be described efficaciously by a single well-defined grouping in a spectral space. Thus, it is better to represent them by a group of spectral classes (clusters), which is prime inference of this method. It is worth noting that this method is identical to the above-mentioned Cluster Pairing (CP) method. In this method, X number of clusters are generated using both live and fake labeled training patches together; we name resulting spectral space as cluster space. For each cluster, an individual classifier (IC) is trained, hence resulting in X number of individual classifiers. Given an image patch, its discriminative value (DV) is computed as an average of the probabilities given by all ICs. Later, the patches are sorted based on their DV with respect to other patches. Finally, patches corresponding to K largest DV values are selected. This method (cluster space + IC) provides a means of optimizing the variance and correlation present in all classes and samples. Following are the steps executed to designate top discriminative patches.

(i) Using training dataset's labeled patches, X number of clusters are generated using both live and fake samples together. K-means clustering algorithm is exploited in this work for cluster creation.

(ii) For each cluster, an individual classifier (IC) (in this work, SVM) is trained using ground truth.

(iii) The patches of a given face image are arranged in descending order on the basis of their respective DV:

$$\text{DV} = \exp\left(\frac{1}{X}\sum_{i=1}^{X} P_i\left(\mathbf{P}\right)\right), \tag{3}$$

where P_i is the probability given by ith classifier trained on ith cluster and $\mathbf{P}$ is the candidate patch.

(iv) The top K patches are selected as discriminative patches.

3.1.2. Nonclustering Based Methods. Unlike the clustering based methods, techniques in this category do not require any clustering algorithm. Following are the four nonclustering based discriminative patch selection methods.

(1) MAXDIST. This method of discriminative patches selection is based on the assumption that candidate discriminative patches are maximally dissimilar from the other patches in the given face image. Therefore, first the dissimilarity between successive patches is computed. The method then sorts the patches based on their average dissimilarity score with respect to other patches and selects those patches (i.e., discriminative patches) that correspond to the K largest average dissimilarity scores. We refer to this method as MAXDIST since discriminative patches are selected using a maximum dissimilarity criterion.

Following steps are followed to select top discriminative patches.

(i) An $n \times n$ dissimilarity matrix ($\mathbf{D}$) is generated, where each element $\mathbf{D}(i, j)$, $i, j \in 1, 2, \ldots, n$ is the distance score between features of patches i and j.

(ii) For the jth patch, the average dissimilarity score (d_j) with respect to the remaining $(n - 1)$ patches is computed by finding the average of the elements in jth row of $\mathbf{D}$.

(iii) The average values obtained in step (ii) are ordered in descending order and the top K patches that have the largest average dissimilarity scores are selected as discriminative patches, since they are the most "dissimilar" patches in the image and hence they are representing typical data measurements.

For classification performance point of view, smaller K values might not be able to sufficiently seize the inter- and intraclass variability, which may lead to inferior performance. However, larger K values, on the other hand, would be computationally demanding. Thus, a rational value of K, by taking above-mentioned factors into account, has to be specified. A similar method has been proposed in [32] for peak frame selection in a given facial expression video. Unlike [32], in this work, we employ the technique to select discriminative patches in an image/frame.

(2) IQA (Image Quality Assessment). This algorithm uses image quality measures to select discriminative patches, thus named as image quality assessment (IQA). The method

assumes that the discriminative patches will have different quality from nondiscriminative patches. The expected quality differences between discriminative and nondiscriminative patches may be local artifacts, color levels, luminance levels, degree of sharpness, entropy, structural distortions, or natural appearance.

This framework exploits 4 general reference-based image quality measures, thus having a very low degree of complexity. First, four distinct label-sets for all patches are obtained using four different image quality measures. The labels are "discriminative" and "nondiscriminative." The method selects only those patches, which are flagged as discriminative by all four-image quality assessment techniques.

In particular, reference-based IQA methods are employed in this scheme that rely on the availability of a clean undistorted reference image to estimate the quality of the test image. However, in typical spoof recognition such a reference image is unknown, because only the input sample is available. Therefore, to circumvent this limitation, the same technique (filtering the image with a low-pass Gaussian kernel) that is successfully being used for image manipulation detection [33] and for steganalysis [34] is utilized. Following steps are executed to attain top K discriminative patches:

(i) The normalized face image ($\mathbf{I}$) is filtered with a low-pass Gaussian kernel in order to generate a smoothed version $\widehat{\mathbf{I}}$.

(ii) Two corresponding sets of dense patches $\{\mathbf{P}_i(\mathbf{I}) \in \mathbb{R}^{M\times N}\}_{i=1}^{n}$ and $\{\mathbf{P}_j(\widehat{\mathbf{I}}) \in \mathbb{R}^{M\times N}\}_{j=1}^{n}$ are taken, where n is the total number of patches.

(iii) Four label matrices ($\mathbf{L}_{\text{PSNR}}$, $\mathbf{L}_{\text{NCC}}$, $\mathbf{L}_{\text{TED}}$, and $\mathbf{L}_{\text{GMSD}}$) using following four-image quality measures (IQM) are generated. The patches are flagged as "discriminative" if their IMQ is greater than or equal to the threshold. The image quality measures are as follows:

(a) *Peak Signal to Noise Ratio (PSNR)*. It computes the distortion between two corresponding patches (of size $P\times Q$) on the basis of their pixel-wise differences as follows:

$$\text{PSNR}\left(\mathbf{P},\widehat{\mathbf{P}}\right) = 10\log\left[\frac{\max\left(\mathbf{P}^2\right)}{\text{MSE}\left(\mathbf{P},\widehat{\mathbf{P}}\right)}\right], \tag{4}$$

where

$$\text{MSE (Mean Squared Error)} = \frac{1}{PQ}\sum_{x=1}^{P}\sum_{y=1}^{Q}\left(\mathbf{P}_{x,y}-\widehat{\mathbf{P}}_{x,y}\right)^2. \tag{5}$$

(b) *Normalized Cross-Correlation (NCC)*. The correlation function can also be used to quantify the similarity between two digital image patches [3]. Here, a variant of correlation based measure is obtained by considering the statistics of the angles between the pixel vectors of the original and distorted patches as

$$\text{NCC}\left(\mathbf{P},\widehat{\mathbf{P}}\right) = \frac{\sum_{x=1}^{P}\sum_{y=1}^{Q}\left(\mathbf{P}_{x,y}\cdot\widehat{\mathbf{P}}_{x,y}\right)}{\sum_{x=1}^{P}\sum_{y=1}^{Q}\left(\mathbf{P}_{x,y}\right)^2}. \tag{6}$$

(c) *Total Edge Difference (TED)*. Edge features are generally considered as one of the most informative datasets in computer vision algorithms. Thus, we considered edge-related quality measure since the structural distortion of an image is deeply linked with its edge degradation. The TED measure is calculated as follows:

$$\text{TED}\left(\mathbf{P},\widehat{\mathbf{P}}\right) = \frac{1}{PQ}\sum_{x=1}^{P}\sum_{y=1}^{Q}\left|\mathbf{P}_{\mathbf{E}_{x,y}}-\widehat{\mathbf{P}}_{\mathbf{E}_{x,y}}\right|. \tag{7}$$

In this work, we use Sobel operator to build the binary edge maps $\mathbf{P}_{\mathbf{E}}$ and $\widehat{\mathbf{P}}_{\mathbf{E}}$.

(d) *Gradient Magnitude Similarity Deviation (GMSD)*. GMSD finds the pixel-wise Gradient Magnitude Similarity (GMS) between the reference and distorted patches. It uses a pooling strategy based on standard deviation of the GMS map to predict accurately perceptual image quality. The GMSD is obtained as follows:

$$\text{GMSD}\left(\mathbf{P},\widehat{\mathbf{P}}\right) = \sqrt{\frac{1}{U}\sum_{i=1}^{U}(\text{GMS}(i)-\text{GMSM})^2}, \tag{8}$$

where U is the total number of pixels in the patch, GMS is Gradient Magnitude Similarity map, and GMSM is Gradient Magnitude Similarity Mean calculated by applying average pooling to the GMS map. The GMS is calculated as

$$\text{GMS}(i) = \frac{2\mathbf{m}_r(i)\,\mathbf{m}_d(i)+c}{\mathbf{m}_r^2(i)+\mathbf{m}_d^2(i)+c}, \tag{9}$$

where c is a positive constant that supplies numerical stability, while $\mathbf{m}_r$ and $\mathbf{m}_d$ are gradient magnitude images obtained by convolution using Prewitt filters along horizontal and vertical directions, respectively. The GMSM is obtained as

$$\text{GMSM} = \frac{1}{U}\sum_{i=1}^{U}\text{GMS}(i). \tag{10}$$

Clearly, a higher GMSM score means higher image quality. We refer reader to [35] for further details of GMSD technique.

(iv) The patches flagged as discriminative by all above four-image quality assessment techniques are selected:

$$\mathbf{L}_{\text{PSNR}}\cap\mathbf{L}_{\text{NCC}}\cap\mathbf{L}_{\text{TED}}\cap\mathbf{L}_{\text{GMSD}}. \tag{11}$$

(v) The patches selected in step (iv) are arranged in descending order on the basis of their average values of (4), (6), (7), and (8).

(vi) The top K patches are selected as discriminative patches.

(3) DF (Diversity Filter). This method considers that the discriminative patches are discernibly diverse from the other patches in the given face image. The discriminative patches are obtained using the combination of a trained classifier (we refer to such classifier as diversity filter) and a selection procedure that selects the patches based on their incremental contribution to the discriminative patch detection performance. This method is close to object detection methods [36, 37], where similar procedure is carried out to capture visually varied parts of objects at a fixed pose or viewpoint. Unlike the proposed method, techniques in [36, 37] use, before diversity filter, a preprocessing step, that is, creating various clusters corresponding to different parts appearances in images. Also distinct diversity filter is employed for each such cluster, while our method single diversity filter is used for all samples and image patches without any clustering. Moreover, the diversity tradeoff parameter in our method is computed dynamically for each image, whereas in [36, 37] a static value is utilized for all images.

In particular, we learn the diversity model of patches based on their properties that can be computed from the filter itself. The intuition is that across image categories good filter exhibits common traits such as low clutter and gradients that are spatially correlated. Therefore, we train a ranking function with the objective to produce the order of diversity quality of patches. The function uses a weight to control tradeoff between estimated rank of a patch and the diversity it adds, which discourages adding patches similar to the ones already selected, even if this patch is highly ranked. Following are the steps required to select K discriminative patches:

(i) Using training dataset, a classifier (diversity filter, here SVM) is trained.

(ii) The patches (n) of a given face image are arranged in descending order on the basis of following equation:

$$\underset{i}{\operatorname{argmax}} \left\{ \widehat{y}_i - \alpha \max_{1,\ldots,t} S_{i,j} \right\}, \tag{12}$$

where $\widehat{y}_i$ is the probability given by diversity filter, $S_{i,j}$ denotes similarity between patches i and j, and $\alpha = \max(\text{eigenvalues}(\text{cov}(S_{i,j})))$ is the diversity tradeoff parameter. It is worth mentioning that in (12) during the first iteration $t = n$ (total number of patches), and then in each successive iteration t is reduced by 1, such that the patch selected in foregoing iteration is removed.

(iii) The top K patches are selected as discriminative patches.

(4) IPI (Intensity-Based Patch of Interest). Local intensity inhomogeneity can be exploited to find the regions, shapes, and edges of similar kind in an image [38]. However, our aim here is to disregard the image patches (regions) with similar features in order to avoid redundancy. Therefore, in this method to determine the discriminative patches, we apply an approach identical to standard statistical background-subtraction approach (which is most commonly used to accost intensity inhomogeneity) [39]. The proposed method does not use any preprocessing step, that is, foreground and background models based on recursive or nonrecursive techniques like in [39]. Following steps are executed to attain K discriminative patches:

(i) A set of dense patches $\{\mathbf{P}_i(\mathbf{I}) \in \mathbb{R}^{M\times N}\}_{i=1}^{n}$ are taken, where n is the total number of patches (of size $P \times Q$).

(ii) A label matrix ($\mathbf{F}_{\text{IPI}}$) is generated using a standard statistical background-subtraction approach:

$$\mathbf{F}_{\text{IPI}} = \begin{cases} \text{Discriminative}, & \text{if } \dfrac{\sum_{x=1}^{P}\sum_{y=1}^{Q} |\mathbf{P}_i(x,y) - \mathbf{P}_i|}{\sigma(\mathbf{P}_i)} > \eta \\ \text{Nondiscriminative}, & \text{otherwise}, \end{cases} \tag{13}$$

where η is threshold, which is estimated using similar procedure as explained above in IQA method.

(iii) The patches flagged as discriminative in step (ii) are arranged in descending order on the basis of their values using (13).

(iv) The top K patches are selected as final discriminative patches.

3.2. Classification Method. For final classification whether the face is genuine or spoof, we used majority-voting based scheme that exploits four well-known classifiers: support vector machine (SVM), Naive-Bayes (NB), Quadratic Discriminant Analysis (QDA), and Ensemble based on AdaBoost algorithm.

4. Experiments

In this section, we evaluate the proposed approach on two different publicly available databases: REPLAY-ATTACK [4] and CASIA-Face Antispoofing Database (FASD) [22].

4.1. Datasets

4.1.1. REPLAY-ATTACK [4]. This dataset is composed of short videos of both real-access and spoofing attempts of 50 different subjects, acquired with a 320 × 240 resolution camera. The datasets were collected under controlled (with a uniform background and artificial lighting) and adverse (with natural illumination and nonuniform background) conditions. The face spoof attacks were created by forging genuine verification attempts of the respective subjects via printed photos, displayed photos/videos on mobile phone's screen, and displayed photos/videos on HD screen.

Table 2: Summary of two databases used in this study.

Database	Number of subjects	Number of videos	Resolution	Attack type
REPLAY-ATTACK [4]	50	(i) 200 live (ii) 1000 spoof	320 × 240	(i) Printed photo (ii) Displayed photo (mobile/HD) (iii) Replayed video (mobile/HD)
CASIA-FASD [22]	50	(i) 150 live (ii) 450 spoof	640 × 480‡ 480 × 640⊕ 1280 × 720⋆	(i) Printed photo (ii) Cut photo (iii) Replayed video

‡, ⊕, and ⋆ indicate low-, normal-, and high-quality camera.

4.1.2. CASIA-FASD [22]. This database contains video recordings of real and fake faces for 50 different identities. Both real-access and spoof attacks were captured using three camera resolutions: low resolution, normal resolution, and high resolution. Three kinds of attack attempts were considered: warped photo attacks, cut photo attacks, and video attacks. The dataset is divided into two subsets for training and testing: 20 and 30 identities, respectively. Table 2 provides a summary of the above two databases.

4.2. Evaluation Protocols. For REPLAY-ATTACK dataset, we followed the same standard protocols specified in [4] for the experiments. The dataset contains three totally independent datasets in terms of users. The train and development sets are used for training and parameter tuning, respectively. The final results are computed on test. The performance of the proposed liveness detection system was evaluated, as in [3, 4, 17], using Half Total Error Rate (HTER) computed as HTER = (FAR + FRR)/2, where FRR and FAR stand for False Rejection Rate and False Acceptance Rate, respectively.

For CASIA-FASD database, we followed the benchmark protocols specified in [22]. The test protocol consists of seven scenarios. The first three scenarios are to study the effect of imaging quality: (1) low quality, (2) normal quality, and (3) high quality. The next three scenarios are (4) warped photo attacks, (5) cut photo attacks, and (6) video attacks. Finally, (7) is the overall scenario (here all data are combined together to give a general and overall evaluation). The classifier training and parameters tuning were performed on training set, while the results are reported in terms of Equal Error Rate (EER) on the test set.

In a given video frame, first the face is detected. The detected face image is then normalized to 128 × 128 pixels and densely divided into a grid of nonoverlapping local patches of size 16 × 16. Out of total n number of patches, only 40% peculiar patches are selected as discriminative patches. The LBP (Local Binary Patterns) technique is utilized to extract the features both for final classification and for discriminative patch selection methods (to obtain dis(similarity) score and clustering). Figure 5 shows examples from REPLAY-ATTACK and CASIA-FASD database of a given face image and corresponding selected discriminative patches using proposed seven image patches selection methods.

4.3. Experimental Results. The experimental results on REPLAY-ATTACK and CASIA-FASD databases are reported in Tables 3 and 4, respectively. We compared the performance of proposed method with most eminent techniques published in the literature such as methodologies in [4] (based on local binary pattern features with SVM classifier), [22] (grounded on multiple difference-of-Gaussian (DoG) filters to extract the high frequency information), [23] (using Multiscale Local Binary Patterns with SVM), [3] (which makes use of general full-reference and nonreference image quality measures), [16] (exploiting correlation between head motion and background that is estimated using optical flow), [28] (encoding information with a Histogram of Oriented Optical Flow (HOOF)), [23] (utilizing both texture and motion estimation along with preprocessing for motion magnification), and [1] (based on image distortion analysis features which is combination of specular reflection, blurriness, chromatic moment, and color diversity properties).

The results in both Tables 3 and 4 show that the proposed method in general achieves better accuracy than existing techniques under specific combination of discriminative patch selection method and classification scheme. For instance, in the experiment using the REPLAY-ATTACK database, it is easy to see in Table 3 that when the proposed framework is implemented using DEND-CLUSTERING-Ensemble or MAXDIST-Ensemble combinations, the Half Total Error Rate (HTER) is 5.00%, which is much lower than method in [25] (i.e., PCA + LBP + SVM (20.50%)). Similarly, in Table 4, we can see that the proposed system achieves error rate better than or similar to the state-of-the-art methods under overall scenario.

The MAXDIST patch selection method achieves better performances on average with the four classification techniques used in this study. Additionally, MAXDIST, CS, and DF patch selection algorithms demonstrate good generalization capability not only for disparate datasets but also for spoofing attacks with varying qualities and fabrication methods. Quite to the contrary, CP and IQA methods fail to attain proficient generalization aptitude. Beside patch selection algorithm, choice of feature classification scheme also plays vital role in accomplishing preferable performances. To this end, it is easy to see in Tables 3 and 4 that, among SVM, Naive-Bayes (NB), QDA, and Ensemble based on AdaBoost classifiers, Ensemble performs best under varying features, datasets, attack types, and amount of training and testing samples, owing to its ability of reducing the variances, averaging out the biases, and most unlikeliness of overfitting. The NB and QDA classifiers in this study are quite sensitive

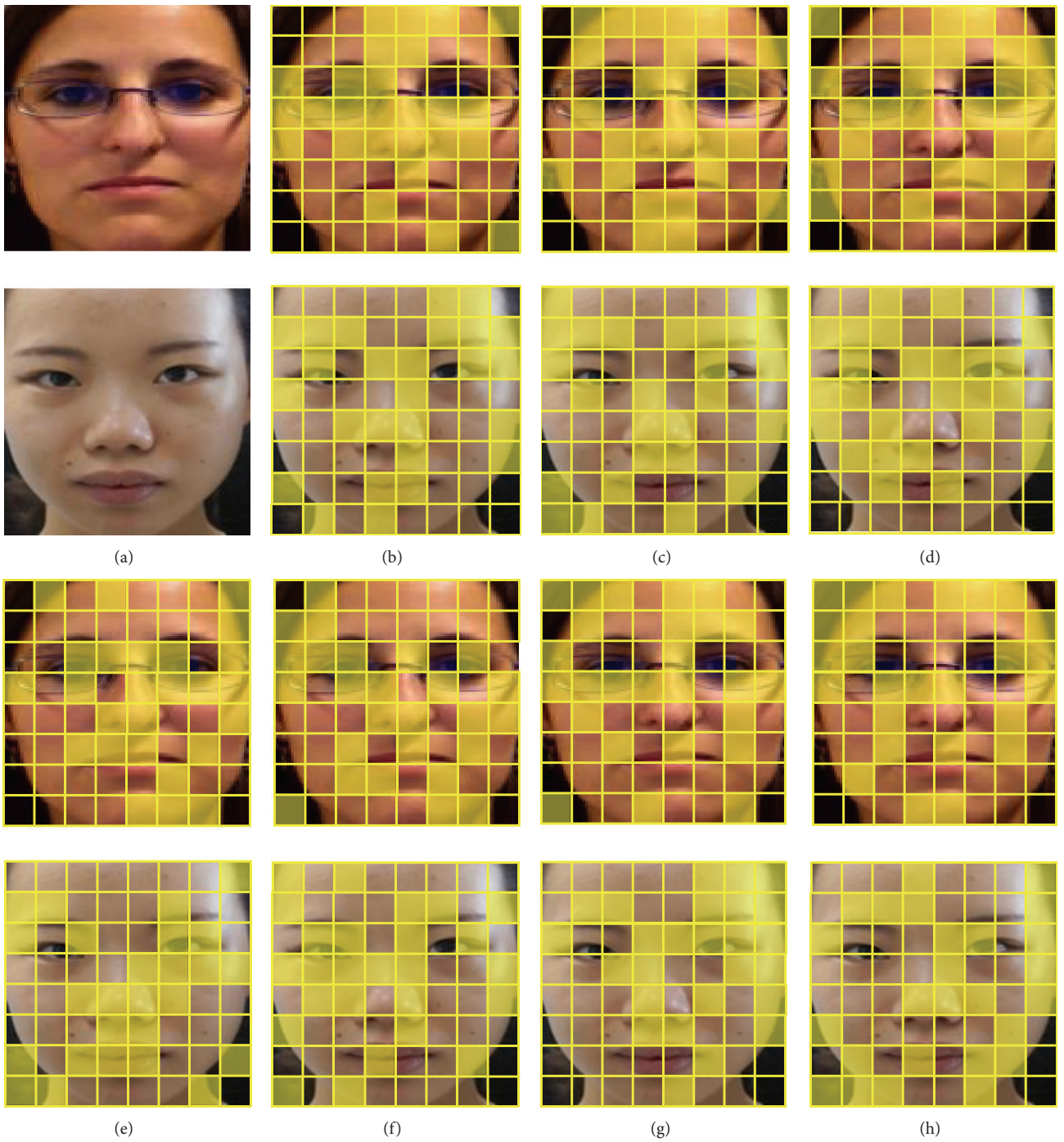

Figure 5: Examples of selected discriminative patches using proposed patches selection methods. Top row of (a)–(h), face from REPLAY-ATTACK database. Bottom row of (a)–(h), face from CASIA-FASD database. A normalized face image (a) of a subject and the discriminative selected patches using (b) CS, (c) DEND-CLUSTERING, (d) IPI, (e) IQA, (f) CP, (g) MAXDIST, and (h) DF methods.

to patch selection approaches. Specifically, though NB enjoys simplicity and computational efficiency, it substantially performs poorly under diverse attack conditions. This may be occurring due to its assumption that all attributes are independent (i.e., no correlation between variables), since it was pointed out in [20, 40] that correlation mapping is beneficial to procure better accuracy and generalization capability in biometric liveness detection. Moreover, NB also assumes that the samples follow Gaussian distribution. However, Gaussian distribution assumption is generally true for small biometric datasets. But, spoofing databases are heterogeneous that contain different spoof attack types and sizes, and thereby NB either gets overfitted or fails to address the problem of concept-drift.

Table 3: Comparison of the proposed method (with SVM, QDA, Naive-Bayes (NB), and Ensemble based classifiers) on REPLAY-ATTACK database with existing methods.

Method	HTER (%)
Multi-LBP [23]	20.25
IQA [3]	15.20
GLCM (Unicamp) [24]	15.62
$LBP^{u2}_{8,1}$ [4]	16.10
$LBP^{u2}_{8,2} + LBP^{u2}_{16,2} + LBP^{u2}_{8,1}$ + SVM [4]	13.87
PCA + LBP + SVM [25]	20.50
Motion [16]	11.70
DoG-LBP + SVM [1]	11.10
LBP-TOP [26]	8.51
IDA [1]	7.41
Proposed: DF-SVM	6.87
Proposed: DF-NB	8.01
Proposed: DF-QDA	7.30
Proposed: DF-Ensemble	6.23
Proposed: CS-SVM	6.25
Proposed: CS-NB	7.44
Proposed: CS-QDA	6.87
Proposed: CS-Ensemble	6.00
Proposed: DEND-CLUSTERING-SVM	5.98
Proposed: DEND-CLUSTERING-NB	8.87
Proposed: DEND-CLUSTERING-QDA	6.11
Proposed: DEND-CLUSTERING-Ensemble	**5.00**
Proposed: IQA-SVM	6.23
Proposed: IQA-NB	11.05
Proposed: IQA-QDA	7.75
Proposed: IQA-Ensemble	5.62
Proposed: IPI-SVM	7.50
Proposed: IPI-NB	8.30
Proposed: IPI-QDA	6.19
Proposed: IPI-Ensemble	6.00
Proposed: CP-SVM	8.37
Proposed: CP-NB	9.18
Proposed: CP-QDA	7.12
Proposed: CP-Ensemble	6.80
Proposed: MAXDIST-SVM	5.87
Proposed: MAXDIST-NB	8.01
Proposed: MAXDIST-QDA	6.12
Proposed: MAXDIST-Ensemble	**5.00**

By metaknowledge analysis, it was found that spoof attacks with higher resolution/quality are comparatively harder to be recognized, as also pointed out in [3, 22], for instance, high-quality eye cut-off attacks in which the cropped portions are filled by real eyes of the attackers leading thus to the high quality spoofing attacks that are having a combination of real and spoofed face features. Furthermore, between REPLAY-ATTACK and CASIA-FASD databases, CASIA-FASD database is more challenging as it incorporates more practical states such as variant of spoof attacks (e.g., cut photo attack simulating eye-blinking) and samples with high quality (resolutions). All in all, results also suggest that, for systems or datasets based on low- or normal-quality samples, it is advisable to adopt CS method with Ensemble classifier to reach desirable performance accuracies, while MAXDIST with Ensemble is better choice for systems/datasets based on high-quality samples, especially videos.

On the whole, it can be stated that use of only certain image locations can significantly influence the face anti-spoofing accuracy. Namely, the proposed method uses only selected discriminative patches and attains higher-ranking precision, unlike the state-of-the-art methods which exploit whole face image/frame/video, leading hence generally to the clutter in the feature representations and to their low performances.

In many face recognition applications, there is no access to the video or image sequences of the user. However, a large number of existing face antispoofing solutions need video or sequences of images (i.e., either for motion or for temporal information) to attain high accuracy. Accordingly, they have less usability, since they are not devised to work on a single static face image. Conversely, the proposed method is single-image algorithm (i.e., the method requires just one input image and not a sequence of them). Therefore, the proposed method is more useful in various applications. Further, it is evident from the experimental results that the proposed framework is robust and performs well across diverse types of spoof attacks, materials, and techniques (such as printed image attack, video-replayed attack, cut photo attack, and image or video resolutions), although only specific face image area is considered. Consequently, the proposed method is much realistic and useful in real-world where a priori attack (artifact) types (paper, mobile, or resolution), which attacker might utilize, are unpredictable.

To sum up, the performances shown by proposed algorithm confirm that contemplating the inherent differences of discriminant abilities among various face image locations is useful for consistently recognizing well the facial spoof attacks. In other words, we show that it seems feasible to use only certain face image patches instead of whole face image to reduce significantly the error rates.

5. Conclusion

The vulnerability of face recognition systems to spoofing attacks is a largely accepted reality, which has led to great advances in face antispoofing (especially face liveness detection) technologies. Despite the remarkable advances, counteracting face spoof attacks has yet proven to be a challenging task. Moreover, existing face liveness detection methods use whole face image or complete video for liveness detection. However, often image regions (video frames) are redundant or correspond to the clutter in the image (video), thus leading generally to low performances. Therefore, in this paper, we propose using just discriminative image patches for face liveness detection. In particular, we present seven

Table 4: Comparison of the proposed method (with SVM, QDA, Naive-Bayes (NB), and Ensemble based classifiers) on CASIA-FASD database with existing methods in terms of EER (%).

Method	Low quality	Normal quality	High quality	Warped photo attack	Cut photo attack	Video attack	Overall
DoG [22]	13.00	13.00	26.00	16.00	6.00	24.00	17.00
IQA [27]	31.70	22.20	5.69	26.10	18.31	34.41	32.45
LBP + SVM baseline	16.50	17.20	23.00	24.70	16.70	27.00	24.80
Multi-LBP baseline	12.77	16.66	26.66	15.55	25.55	17.77	17.77
HOOF [28]	16.66	30.00	26.11	15.55	17.77	38.88	21.11
Mag-HOOF	17.22	33.33	22.77	12.22	20.00	36.60	22.22
HOOF + Multi-LBP	9.44	20.55	16.66	10.00	16.66	24.44	15.55
Motion-MLBP [23]	7.22	13.33	29.44	14.44	22.22	13.33	15.74
Motion magnification [23]	6.11	23.33	13.88	10.00	14.44	20.00	14.44
Color texture [29]	7.80	10.10	6.40	7.50	**5.40**	8.10	6.20
Proposed: DF-SVM	7.53	6.65	6.28	6.94	8.21	7.97	6.71
Proposed: DF-NB	7.77	7.79	6.66	7.00	7.66	8.16	9.00
Proposed: DF-QDA	5.78	7.01	5.65	6.97	7.88	7.15	7.81
Proposed: DF-Ensemble	**4.65**	5.99	6.57	5.94	6.49	6.00	6.11
Proposed: CS-SVM	6.46	6.43	5.97	6.56	8.72	7.27	8.54
Proposed: CS-NB	6.69	6.30	5.99	7.49	8.01	7.68	8.69
Proposed: CS-QDA	7.33	8.89	6.09	7.43	8.21	7.68	7.97
Proposed: CS-Ensemble	6.51	**5.59**	**5.30**	**5.31**	6.13	7.12	7.59
Proposed: DEND-CLUSTERING-SVM	7.39	7.09	5.93	7.35	8.22	8.42	8.07
Proposed: DEND-CLUSTERING-NB	6.98	6.72	7.40	7.65	8.23	8.00	8.45
Proposed: DEND-CLUSTERING-QDA	7.11	6.89	7.90	7.03	6.99	8.89	8.67
Proposed: DEND-CLUSTERING-Ensemble	5.89	6.06	5.58	5.33	5.42	6.02	5.16
Proposed: IQA-SVM	6.32	7.00	5.97	7.73	7.71	7.69	7.33
Proposed: IQA-NB	6.66	7.66	6.32	7.33	8.45	8.00	8.22
Proposed: IQA-QDA	6.35	7.78	8.31	8.78	7.87	8.06	8.69
Proposed: IQA-Ensemble	6.02	6.57	5.76	7.27	6.87	7.27	7.27
Proposed: IPI-SVM	8.34	7.99	8.72	7.89	8.56	8.74	8.36
Proposed: IPI-NB	8.88	7.45	8.35	7.22	8.90	8.12	8.83
Proposed: IPI-QDA	6.01	7.33	6.88	7.33	8.78	9.14	8.57
Proposed: IPI-Ensemble	6.00	6.78	6.10	6.27	7.24	8.00	7.22
Proposed: CP-SVM	8.87	8.23	9.08	7.64	9.37	8.90	9.50
Proposed: CP-NB	9.10	9.07	8.39	9.74	9.85	9.45	9.34
Proposed: CP-QDA	8.67	8.34	9.00	8.78	8.51	8.34	8.59
Proposed: CP-Ensemble	8.01	7.27	7.34	6.82	7.13	8.23	7.60
Proposed: MAXDIST-SVM	7.25	5.76	6.68	9.28	8.28	8.42	8.56
Proposed: MAXDIST-NB	7.37	6.98	7.33	7.33	8.00	8.31	8.00
Proposed: MAXDIST-QDA	7.11	6.78	7.32	8.84	8.02	8.42	8.40
Proposed: MAXDIST-Ensemble	5.26	6.00	**5.30**	5.78	5.49	**5.02**	**5.07**

novel methods to obtain discriminative patches in a face image (or randomly selected lone video frame). The features of selected discriminative image patches are fed to a specific classifier (i.e., SVM, Naive-Bayes, QDA, or Ensemble). The classification results of these patches are combined by a majority-voting based scheme for the final classification of genuine and spoof faces. Experimental results on two publicly available databases show comparative performances compared to the existing works. The future works include devising more novel techniques for attaining discriminative image patches and inclusion of temporal information in the proposed method for higher security applications.

Competing Interests

The authors declare that they have no competing interests.

References

[1] 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.

[2] Z. Akhtar, C. Micheloni, and G. L. Foresti, "Biometric liveness detection: challenges and research opportunities," *IEEE Security & Privacy*, vol. 13, no. 5, pp. 63–72, 2015.

[3] J. Galbally, S. Marcel, and J. Fierrez, "Image quality assessment for fake biometric detection: application to iris, fingerprint, and face recognition," *IEEE Transactions on Image Processing*, vol. 23, no. 2, pp. 710–724, 2014.

[4] I. Chingovska, A. Anjos, and S. Marcel, "On the effectiveness of local binary patterns in face anti-spoofing," in *Proceedings of the International Conference of the Biometrics Special Interest Group (BIOSIG '12)*, pp. 1–7, Darmstadt, Germany, September 2012.

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

[6] Z. Akhtar, C. Micheloni, C. Piciarelli, and G. L. Foresti, "MoBio_LivDet: mobile biometric liveness detection," in *Proceedings of the 11th IEEE International Conference on Advanced Video and Signal Based Surveillance (AVSS '14)*, pp. 187–192, Seoul, Republic of Korea, August 2014.

[7] G. Pan, L. Sun, Z. Wu, and S. Lao, "Eyeblink-based anti-spoofing in face recognition from a generic webcamera," in *Proceedings of the IEEE 11th International Conference on Computer Vision (ICCV '07)*, pp. 1–8, Rio de Janeiro, Brazil, October 2007.

[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 *Proceedings of the 11th European Conference on Computer Vision (ECCV '10)*, K. Daniilidis, P. Maragos, and N. Paragios, Eds., vol. 6316 of *Lecture Notes in Computer Science*, pp. 504–517, Crete, Greece, September 2010.

[9] Y. Li and X. Tan, "An anti-photo spoof method in face recognition based on the analysis of fourier spectra with sparse logistic regression," in *Proceedings of the Chinese Conference on Pattern Recognition (CCPR '09)*, Nanjing, China, November 2009.

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

[11] W. Bao, H. Li, N. Li, and W. Jiang, "A liveness detection method for face recognition based on optical flow field," in *Proceedings of the International Conference on Image Analysis and Signal Processing (IASP '09)*, pp. 233–236, IEEE, Taizhou, China, April 2009.

[12] G. Pan, L. Sun, Z. Wu, and Y. Wang, "Monocular camera-based face liveness detection by combining eyeblink and scene context," *Telecommunication Systems*, vol. 47, no. 3, pp. 215–225, 2011.

[13] J. Li, Y. Wang, T. Tan, and A. K. Jain, "Live face detection based on the analysis of Fourier spectra," in *Proceedings of the Biometric Technology for Human Identification*, vol. 5404 of *Proceedings of SPIE*, pp. 296–303, Orlando, Fla, USA, April 2004.

[14] J. Bai, T. Ng, X. Gao, and Y. Shi, "Is physics-based liveness detection truly possible with a single image?" in *Proceedings of IEEE International Symposium on Circuits and Systems*, pp. 3425–3428, Paris, France, May-June 2010.

[15] Z. Zhang, D. Yi, Z. Lei, and S. Z. Li, "Face liveness detection by learning multispectral reflectance distributions," in *Proceedings of the IEEE International Conference on Automatic Face and Gesture Recognition and Workshops (FG '11)*, pp. 436–441, Santa Barbara, Calif, USA, March 2011.

[16] J. Komulainen, A. Hadid, M. Pietikäinen, A. Anjos, and S. Marcel, "Complementary countermeasures for detecting scenic face spoofing attacks," in *Proceedings of the International Conference on Biometrics (ICB '13)*, pp. 1–7, Madrid, Spain, June 2013.

[17] A. Anjos and S. Marcel, "Counter-measures to photo attacks in face recognition: a public database and a baseline," in *Proceedings of the International Joint Conference on Biometrics (IJCB '11)*, pp. 1–7, Washington, DC, USA, October 2011.

[18] T. Wang and S. Z. Li, "Face liveness detection using 3d structure recovered from a single camera," in *Proceedings of the International Conference on Biometrics (ICB '13)*, pp. 1–6, IEEE, Madrid, Spain, 2013.

[19] T. I. Dhamecha, A. Nigam, R. Singh, and M. Vatsa, "Disguise detection and face recognition in visible and thermal spectrums," in *Proceedings of the 6th IAPR International Conference on Biometrics (ICB '13)*, pp. 1–6, Madrid, Spain, June 2013.

[20] G. Chetty and M. Wagner, "Liveness detection using cross-modal correlations in face-voice person authentication," in *Proceedings of the 9th European Conference on Speech Communication and Technology (Interspeech '05)*, pp. 2181–2184, Lisbon, Portugal, September 2005.

[21] N. Kose and J.-L. Dugelay, "Reflectance analysis based countermeasure technique to detect face mask attacks," in *Proceedings of the 18th International Conference on Digital Signal Processing (DSP '13)*, pp. 1–6, IEEE, Fira, Greece, July 2013.

[22] Z. Zhang, J. Yan, S. Liu, Z. Lei, D. Yi, and S. Z. Li, "A face antispoofing database with diverse attacks," in *Proceedings of the 5th IAPR International Conference on Biometrics (ICB '12)*, pp. 26–31, IEEE, New Delhi, India, April 2012.

[23] S. Bharadwaj, T. I. Dhamecha, M. Vatsa, and R. Singh, "Face anti-spoofing via motion magnification and multifeature videolet aggregation," Tech. Rep. IIITD-TR-2014-002, 2014.

[24] I. Chingovska, J. Yang, Z. Lei, and D. Yi, "The 2nd competition on counter measures to 2D face spoofing attacks," in *Proceedings of the IEEE International Conference on Biometrics (ICB '13)*, pp. 1–6, Madrid, Spain, June 2013.

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

[26] 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 *Proceedings of the 6th IAPR International Conference on Biometrics (ICB '13)*, pp. 1–8, Madrid, Spain, June 2013.

[27] J. Galbally and S. Marcel, "Face anti-spoofing based on general image quality assessment," in *Proceedings of the 22nd International Conference on Pattern Recognition (ICPR '14)*, pp. 1173–1178, Stockholm, Sweden, August 2014.

[28] R. Chaudhry, A. Ravichandran, G. Hager, and R. Vidal, "Histograms of oriented optical flow and Binet-Cauchy kernels on nonlinear dynamical systems for the recognition of human actions," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR '09)*, pp. 1932–1939, Miami, Fla, USA, June 2009.

[29] Z. Boulkenafet, J. Komulainen, and A. Hadid, "Face anti-spoofing based on color texture analysis," in *Proceedings of the IEEE International Conference on Image Processing (ICIP '15)*, pp. 2636–2640, Quebec City, Canada, September 2015.

[30] M. Nilsson, J. Nordberg, and I. Claesson, "Face detection using local SMQT features and split up snow classifier," in *Proceedings of the IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP '07)*, pp. II-589–II-592, Honolulu, Hawaii, USA, April 2007.

[31] U. Uludag, A. Ross, and A. K. Jain, "Biometric template selection and update: a case study in fingerprints," *Pattern Recognition*, vol. 37, no. 7, pp. 1533–1542, 2004.

[32] S. Zhalehpour, Z. Akhtar, and C. Eroglu Erdem, "Multimodal emotion recognition based on peak frame selection from video," *Signal, Image and Video Processing*, 2015.

[33] S. Bayram, I. Avcibaş, B. Sankur, and N. Memon, "Image manipulation detection," *Journal of Electronic Imaging*, vol. 15, no. 4, Article ID 041102, 17 pages, 2006.

[34] I. Avcibas, N. Memon, and B. Sankur, "Steganalysis using image quality metrics," *IEEE Transactions on Image Processing*, vol. 12, no. 2, pp. 221–229, 2003.

[35] 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.

[36] L. Bourdev and J. Malik, "Poselets: body part detectors trained using 3D human pose annotations," in *Proceedings of the IEEE 12th International Conference on Computer Vision (ICCV '09)*, pp. 1365–1372, Kyoto, Japan, September 2009.

[37] P. F. Felzenszwalb, R. B. Girshick, D. McAllester, and D. Ramanan, "Object detection with discriminatively trained part-based models," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 32, no. 9, pp. 1627–1645, 2010.

[38] U. Vovk, F. Pernus, and B. Likar, "A review of methods for correction of intensity inhomogeneity in MRI," *IEEE Transactions on Medical Imaging*, vol. 26, no. 3, pp. 405–421, 2007.

[39] S. Cheung and C. Kamath, "Robust techniques for background subtraction in urban traffic video," in *Proceedings of the IEEE Conference on Visual Communications and Image Processing (VCIP '07)*, pp. 1–12, 2007.

[40] Z. Akhtar, C. Micheloni, and G. L. Foresti, "Correlation based fingerprint liveness detection," in *Proceedings of the International Conference on Biometrics (ICB '15)*, pp. 305–310, Phuket City, Thailand, May 2015.