IEEE Access
Multidisciplinary | Rapid Review | Open Access Journal

Received May 29, 2017, accepted July 1, 2017, date of publication July 19, 2017, date of current version August 8, 2017.

Digital Object Identifier 10.1109/ACCESS.2017.2729161

An Anomaly Detection Approach to Face Spoofing Detection: A New Formulation and Evaluation Protocol

SHERVIN RAHIMZADEH ARASHLOO[1], JOSEF KITTLER[2], (Life Member, IEEE), AND WILLIAM CHRISTMAS, (Member, IEEE)

[1]Department of Medical Informatics, Faculty of Medical Sciences, Tarbiat Modares University, Tehran 111-14115, Iran
[2]Centre for Vision, Speech and Signal Processing, University of Surrey, Guildford GU2 7XH, U.K.

Corresponding author: Shervin Rahimzadeh Arashloo (s.rahimzadeh@modares.ac.ir)

This work was supported in part by EPSRC under Project EP/K014307/1 and Project EP/N007743/1 and in part by the European Union Project Beat.

ABSTRACT Face spoofing detection is commonly formulated as a two-class recognition problem where relevant features of both positive (real access) and negative samples (spoofing attempts) are utilized to train the system. However, the diversity of spoofing attacks, any new means of spoofing attackers, may invent (previously unseen by the system) the problem of imaging sensor interoperability, and other environmental factors in addition to the small sample size make the problem quite challenging. Considering these observations, in this paper, a number of propositions in the evaluation scenario, problem formulation, and solving are presented. First of all, a new evaluation protocol to study the effect of occurrence of unseen attack types, where the train and test data are produced by different means, is proposed. The new evaluation protocol better reflects the realistic conditions in spoofing attempts where an attacker may come up with new means for spoofing. Inter-database and intra-database experiments are incorporated into the evaluation scheme to account for the sensor interoperability problem. Second, a new and more realistic formulation of the spoofing detection problem based on the anomaly detection concept is proposed where the training data come from the positive class only. The test data, of course, may come from the positive or negative class. Such a one-class formulation circumvents the need for the availability of negative training samples, which, in an in deal case, should be the representative of all possible spoofing types. Finally, a thorough evaluation and comparison of 20 different one-class and two-class systems on the video sequences of three widely employed databases is performed to investigate the merits of the one-class anomaly detection approaches compared with the common two-class formulations. It is demonstrated that the anomaly-based formulation is not inferior as compared with the conventional two-class approach.

INDEX TERMS Face spoofing detection, anomaly detection, one-class classification, inter-type face spoofing detection.

I. INTRODUCTION

Spoofing attacks are known to pose a challenge to the deployment of biometrics systems and face recognition algorithms are no exception. In a spoofing scenario, an imposter tries to get illegal access to a service by presenting artificial biometrics traits of a registered subject. These typically include print attacks, and replay attacks.

The common approach to detecting spoofing attacks is to collect both real and fake data (spoofing attempts) and then try to learn a suitable two-class classifier to predict whether a test sample is a real access or a spoofing attempt. However, despite the great progress made in this direction [1]–[5], there are certain drawbacks to this approach. First of all, while all the real-access data are assumed to be ideally of the same nature, the spoofing attacks can potentially be very diverse. In other words, the data which represent the negative class (spoofing attack) does not necessarily form a compact region in the chosen representation space. This can be due to, for example, different media used for spoofing attempts such as printed photo attack or video replay attack. The resulting multimodality of spoofing attack data hinders learning an effective decision boundary between the positive

and negative samples. Second, augmenting the training set with more samples to improve classification performance in such two-class schemes is not easy as both real access and fake data are required. In particular, collecting data corresponding to spoofing attacks is labour intensive and hence limited in number. On the other hand, increasing the number of samples only in the positive set would result in imbalanced training data and may not be as effective as expected in improving the classification performance. Third, attackers may come up with new and inventive ways of spoofing which were not previously seen during the training phase. Thus, regardless of the database used for training, the negative training samples may not be representative of all potential spoofing attacks. As a result, the learned two-class classifiers in these situations do not always appropriately generalise to accommodate such unseen test data in the operation phase. Therefore, frequent spoofing signature updates are required. Moreover, most of the time, one is not interested in identifying the source of a spoofing attack (e.g. whether the attacker has used a printed photograph or a video screen to spoof a face recognition system is irrelevant in the current context) but merely detecting its occurrence. The diversity of imaging sensors and the sensor inter-operability problem and other environmental conditions when added to the aforementioned undesirable factors, pose serious challenges in detection of spoofing attempts.

Motivated by the aforementioned observations, the current work proposes new approaches to both detection and system performance evaluation. Regarding the detection mechanism, in the current work, face spoofing detection is formulated as an anomaly detection problem. Anomaly detection refers to finding a pattern exhibiting abnormal properties or behaviour [6]. In the context of the current study, normal behaviour/data corresponds to real access attempts while spoofing attacks are considered as anomalies. The history of anomaly detection can be traced back to the statistics community in the nineteenth century [7]. Since these early efforts, detecting anomalies or outliers has been of interest in many different domains resulting in a variety of anomaly detection mechanisms being suggested over the time. It finds applications in a wide spectrum of domains such as intrusion detection [8], processing astronomical data [9], satellite imagery [10], digit recognition [11], spectroscopy [12], mammographic image analysis [13], video surveillance [14], etc. Anomalies have been referred to as outliers, exceptions, surprises, contaminations, etc. in different domains. More recently, this concept has been extended to domain anomaly, referring to a situation when none of the models characterising a domain appropriately explain the observed data [15]. The concept of anomaly detection has been used for biometric identification [16]–[19] but no other work has performed a thorough investigation on its deployment in the context of face spoofing detection problem.

Representation and modelling normal behaviour partly depends on the training data available. It may be based on data samples representing the normal class, or both on the normal data and samples of abnormal (anomaly) observations. There exist also methods in the literature which focus solely on the anomaly instances for training. However, as it is difficult to construct a training set which covers every possible type of anomaly, such techniques are not commonly deployed. Anomaly detection methods may be broadly classified into one of two major groups: generative and non-generative [15]. While generative methods usually model the distribution functions of observations, the non-generative methods very often are based on classifying normal observations by learning the boundary delineating normality.

The current work formulates face anti-spoofing as a one-class pattern recognition problem where only normal data is used for training. Compared to the existing two-class approaches for face spoofing detection, the proposed one-class framework has appealing characteristics as follows. First, as only the positive (normal) samples are used for building the model, the undesirable effects of the diversity of spoofing data on detection performance is minimised. This is based on the assumption that all the real access (normal) data are more or less similar in nature. Second, in order to improve system performance one may more easily extend the training data. This is due to the fact that only real-access data are required for training. Third, an anomaly detection-based system potentially has the ability to detect previously unknown attacks due to the fact that the constructed model in such an approach is not based on specific signatures representing spoofing attempts. Rather, the system would flag an observation as anomaly (spoofing attempt) only when it deviates from the normal data (real access attempts) and not because the system has faced a known spoofing attack signature.

In terms of the evaluation protocol, the current work proposes a more realistic approach for performance evaluation. Currently, the existing databases and the associated evaluation protocols do not have the potential to accurately predict system behaviour in realistic conditions as a result of at least two disturbing factors. The first deficiency results from the nature of the anomalies for which a globally comprehensive model may be very difficult, if not impossible, to construct. This owes to the fact that different subjects may come up with new and inventive ways of producing anomalous data for spoofing attempts which were not previously captured by the negative samples of the database used in the training phase. This particular aspect of evaluation is quite important in order to construct a reliable system for practical usage which is missing in the evaluation schemes of the current data sets. The second drawback of the existing evaluation protocols is the lack of accounting for sensor interoperability and some other environmental changes in spoofing detection. The representations used for classification purposes may vary undesirably as a result of using different imaging sensors and other undesired variations in imaging conditions during the training and operational phases which may lead to deterioration in system performance. In this respect, in the new evaluation protocol, both inter-database and intra-database evaluations

are incorporated to take into account the variability in imaging sensor and some other environmental conditions. The new evaluation protocol proposed in this work addresses the aforementioned concerns, to be discussed in the subsequent sections.

For the assessment of the proposed anomaly-based spoofing detection mechanism under the new evaluation protocol a thorough examination and comparison of 20 different one-class and two-class systems is performed. The experiments are particularly designed to investigate the merits of the one-class anomaly detection approaches compared with two-class formulations in terms of the generalisation of the two group of methods across different spoofing types as well as across different databases. The experiments are conducted on video sequences of three widely employed databases using common spatio-temporal as well as image quality features. As a by-product of the current work, a comparative study of the commonly used representations for spoofing detection is also performed.

The main contributions of the current work can be summarised as follows. First, a new and more realistic formulation of the face spoofing detection problem, considering spoofing detection as an anomaly detection problem is proposed. Although the methodology has been examined in the face spoofing detection paradigm, nevertheless, it is general and may be equally applicable to other biometric modalities. Second, the one-class pattern recognition problem is solved using standard methods. It is shown that anomaly based formulation is not inferior in comparison with the common two-class formulation. Third, a new evaluation protocol for studying the generalisation performance of the proposed anomaly detection approach as well as the conventional two-class methods is proposed. In this respect, the spoofing detection rates of various systems are examined and compared in different intra-database as well as inter-database and cross-type settings. Forth, the effect of using background information on detection performance in intra- and inter-database settings is examined. Fifth, it is shown that the performance of both one-class and two-class formulations is not adequate and more research is needed to enhance detection rates. And last but not least, as a by-product of the experimental evaluations, a comparison of different representations used for face spoofing detection on video sequences of three publicly available data sets is provided. The rest of the paper is organised as follows. Section II reviews the literature on existing face spoofing detection methods. In Section III, once a description of the employed representations is provided, the anomaly detection mechanisms examined are presented. In Section IV, after giving a brief description of the databases used, the new inter- and intra-database cross-type evaluation settings are presented. The discussion is then followed by an overview of the one-class and two-class systems evaluated and a discussion of the results of their extensive evaluation on the three benchmark databases in various scenarios. Finally, conclusions are drawn in Section V.

II. PRIOR WORK

The existing approaches to spoofing detection operate by training the system using both real access and spoofing data. The differences largely lie in the representations used for modelling real vs. spoofing attempts. Categorisation of existing approaches based on the employed cues has been performed in a recent study [1] where they are broadly classified into three major groups.

The first group comprises methods based on detecting different signs of vitality. Such methods employ characteristics corresponding to live faces. For example, [20] uses blinking for spoofing detection. Eye-blink along with other cues are used in other works. For instance, the work in [21] proposes a hybrid face liveness detection mechanism utilising both eye-blink and scene context clues. Reference [2] advocates the use of all dynamic information content of the video, such as blinking eyes, moving lips, and facial dynamics, which is represented using a dynamic mode decomposition method.

The second group of methods are those which are based on the differences in motion patterns between real and spoofing attacks. These approaches mainly rely on the fact that the spoofing media are often flat 2D planes whereas real accesses correspond to 3D structures. Moreover, it has been assumed that motion in spoofing attacks is rigid whereas both rigid and non-rigid motion is present in real-access attempts. The work in [22] is a typical instance of the methods in this class. It uses Eulerian motion magnification to enhance facial expressions. The method deploys two sets of features composed of LBPs as well as motion cues. Geometric invariants are used for detecting replay attacks once a set of automatically located facial points are detected in [23]. The proposed system was evaluated on two publicly available databases of NUAA [24] and HONDA [25]. Differences in optical flow fields generated by the movements of 2D planes and 3D objects motivated the work in [26] where a new liveness detection method is proposed. Another study [27] proposed a face spoofing countermeasure based on foreground/background motion correlation using optical flow. The method was shown to obtain promising results on the Photo-Attack database.

The last category of face liveness detection methods relies on image distortion/quality measures. As an example of the methods in this category, the work in [28] detects print-attacks using the differences in the 2D Fourier spectra. The study in [29] uses a set of difference-of-Gaussian filters to choose specific frequency bands to be used for feature extraction. The work in [24] uses the Lambertian model and puts forth two methods. The first one is a variational Retinex-based approach while the second uses Gaussian filters difference. Mtt *et al.* [5] propose to detect spoofing via texture analysis. Texture in this work is represented using multi-scale local binary patterns. The work in [30] exploits both frequency and texture information using power spectrum and LBPs. Both spatial and temporal information are modelled in [31] for face anti-spoofing. The descriptors used encoded information about shape, colour and texture. The combination

of motion and texture methods via score level fusion is proposed in [3]. Good performance has been reported on the Replay-Attack database [19].

Some other works require special hardware during image capture for spoofing detection. The work in [32] is one such method illuminating the face during image capture and using the reflected colour from the face as a means of image watermarking. The assumption here is that a pre-recorded video is highly unlikely to contain the correctly reflected colour sequence. In [33], image distortion is modelled as four different features of specular reflection, blurriness, chromatic moment, and colour diversity. Multiple SVMs trained for different face spoofing attacks are used to discriminate between genuine faces and spoofing attacks. The work in [34] investigates suitable convolutional network architectures and tries to learn the weights of the network. In [35] the detection of spoofing is attempted via a combination of LPQ-TOP [36], [37] and BSIF-TOP [38] features. The fusion of multiple sources of information is accomplished via a fast kernel discriminant analysis approach using spectral regression.

While most of the existing methods try to learn a general classifier to delineate spoofing attempts from real access data, the work in [4] proposes a method using a classifier specifically trained for each subject. In a similar attempt, the work in [39] studied the client-specific information embedded in the feature space and its effects on the performance of the system. Using both texture and motion cues, the authors built two anti-spoofing methods, one being a generative approach while the other being a discriminative method.

In terms of detection mechanism, the current work, while sharing some similarities to the existing approaches in the form of image/video content representation, is distinct in the way the detection problem is formulated. Whereas the common approach is to separate the negative from positive samples using a two-class formulation, our proposition is to identify spoofing attempts by means of anomaly detection based on one-class pattern classification methods. Moreover, the evaluations are performed in a newly defined setting which betters reflects the difficulties of detection in realistic scenarios.

III. METHODOLOGY

The current work approaches face spoofing detection problem from an anomaly detection perspective. For this purpose, different one-class models are used to detect anomalous patterns (outliers). The assumption here is that spoofing attempts can potentially be much more diverse in the chosen representation space compared to normal (real access) data instances. A data instance is labelled as an anomaly if it does not belong to the normal class representing real accesses. In this work, we make use of some the most commonly employed descriptors for face spoofing detection. In the remainder of this section, we briefly review the descriptors adopted for representing an image sequence and then discuss the anomaly detection mechanisms proposed in this work.

A. REPRESENTATIONS USED

In order to model the temporal as well as spatial attributes of an image sequence, dynamic texture descriptors are employed in the current work. For this purpose, an image sequence is modelled in three different ways: using a local binary pattern operator on three orthogonal planes (LBP-TOP) [40]; local phase quantization on three orthogonal planes (LPQ-TOP) [36], [37]; and binarised statistical image features on three orthogonal planes (BSIF-TOP) [38]. This choice is driven by their success in face spoofing detection. Moreover, motivated by the success of image quality measures, we use a fourth set of features, consisting of image quality measures employed in [41].

1) LBP-TOP

The local binary pattern (LBP) coding was first introduced in [42]. Assume $I(x, y)$ denotes a gray scale image and g_c represents the value of a pixel at (x_c, y_c) coordinates. In addition, let g_m denote the value of a sampling point in an equally spaced circular neighborhood of radius R which is comprised of P points. The LBP code at (x_c, y_c) is then defined as

$$h_{P,R}(x_c, y_c) = \sum_{m=0}^{P-1} \left(L\left((g_m - g_c) \geq 0\right)\right) 2^m \qquad (1)$$

where $L(x)$ denotes a function with $L(x) = 1$, if x is true, and $L(x) = 0$, otherwise. The occurence of codes is then summarised using histograms. The original LBP operator was proposed to operate on static images. Later, this approach was extended to the spatio-temporal domain applicable to image sequences [40], dubbed LBP on three orthogonal planes (LBP-TOP). The LBP-TOP approach applies the LBP operator separately on three orthogonal planes (XY, XT and YT). The histograms thus obtained then describe appearance, horizontal motion and vertical motion in an image sequence. In the LBP-TOP approach, it is possible to change the radii in axes X, Y and T (R_x, R_y and R_t) as well as the numbers of neighboring points in the XY, XT and YT planes (P_{XY}, P_{XT} and P_{YT}). By varying the radius, a multi-scale representation of time-varying texture can be obtained. In order to acquire a coherent description for videos of different spatial and temporal sizes, the histograms are normalized using L_1-normalization. Finally, a description of a video is obtained by concatenating histograms of different planes into one histogram. The neighbourhood parameters encoding the number of neighbours and their distance to the centre pixel in this work are set to $P_{XY} = P_{XT} = P_{YT} = 8$ and $R_{XY} = R_{XT} = R_{YT} = 1, \ldots, 6$. The LBP-TOP representation has been shown to be effective for face spoofing detection in [1].

2) LPQ-TOP

Motivated by the blur-tolerant property of the Fourier phase spectrum, the local phase quantization (LPQ) [43] uses local phase information obtained by a short-term Fourier transform over a square region for texture description. The short-term Fourier transform over a region of $l \times l$ pixels centered at

pixel position $\mathbf{m}$ of the image $X(\mathbf{y})$ is defined as

$$F(\mathbf{f}, \mathbf{m}) = \sum_{\mathbf{y}} X(\mathbf{m} - \mathbf{y}) e^{-\mathrm{j}2\pi \mathbf{f}^\top \mathbf{y}} = \mathbf{w}_\mathbf{f}^\top \mathbf{x} \quad (2)$$

where $\mathbf{w_f}$ denotes the basis vector of the 2D discrete Fourier transform at frequency $\mathbf{f}$ while $\mathbf{x}$ stands for the vector containing all l^2 pixels of the region in consideration. In the LPQ approach, the local Fourier coefficients are computed at four frequencies $\mathbf{f}_1 = [a, 0]^\top$, $\mathbf{f}_2 = [0, a]^\top$, $\mathbf{f}_3 = [a, a]^\top$ and $\mathbf{f}_4 = [a, -a]^\top$, where a is a suitably chosen small scalar. For each pixel location the result is conveyed as the vector $F^C_\mathbf{m} = [F(\mathbf{f}_1, \mathbf{m}), F(\mathbf{f}_2, \mathbf{m}), F(\mathbf{f}_3, \mathbf{m}), F(\mathbf{f}_4, \mathbf{m})]$. Assume

$$F_\mathbf{m} = [Re\{F^C_\mathbf{m}\}, Im\{F^C_\mathbf{m}\}] \quad (3)$$

where $Re\{.\}$ and $Im\{.\}$ return the real and imaginary parts of a complex number, respectively. The corresponding $8 \times l^2$ transformation matrix which would produce $F_\mathbf{m}$ as $F_\mathbf{m} = T\mathbf{x}$ is then

$$T = [Re\{\mathbf{w}_{\mathbf{f}_1}, \mathbf{w}_{\mathbf{f}_2}, \mathbf{w}_{\mathbf{f}_3}, \mathbf{w}_{\mathbf{f}_4}\}, Im\{\mathbf{w}_{\mathbf{f}_1}, \mathbf{w}_{\mathbf{f}_2}, \mathbf{w}_{\mathbf{f}_3}, \mathbf{w}_{\mathbf{f}_4}\}]^\top \quad (4)$$

The vector $F_\mathbf{m}$ is de-correlated in the next step using a whitening transform to produce uncorrelated vector $G_\mathbf{m}$. Next, the information in the Fourier coefficients is recorded by binarizing the elements of $G_\mathbf{m}$ as

$$q_j = \begin{cases} 1 & \text{if } g_j \geq 0, \\ 0 & \text{otherwise.} \end{cases} \quad (5)$$

where g_j is the j^{th} component of vector $G_\mathbf{m}$. The resultant 8-bit binary coefficients q_j are finally represented as integers and summarised using a histogram representation.

The same approach as in the LBP-TOP method is employed to extend the LPQ descriptor to the spatio-temporal domain [36], [37]. For this purpose, the basic LPQ features are extracted independently from three orthogonal planes: XY, XT and YT and stacked into a single histogram. In order to capture texture details at various scales, the window sizes may be varied. In this work, six different window sizes of $\{3 \times 3, 5 \times 5, \ldots, 13 \times 13\}$ are employed to obtain a multi-scale representation of the dynamic textural content of an image sequence. The histograms of each plane at each scale are L_1-normalized independently to yield a coherent representation and then concatenated to form the final multi-scale LPQ-TOP descriptor. The LPQ-TOP dynamic texture descriptor in conjunction with other representations has been successfully used for face spoofing detection in previous work [35].

3) BSIF-TOP

Binarized statistical image features are based on independent component analysis [44]. The BISF descriptor uses learned filters to extract features from local image patches. Considering an image patch X of size $l \times l$ pixels and a linear filter W_i of a corresponding size, the filter response s_i is obtained by convolution as

$$s_i = \sum_{\mathbf{y}} W_i(\mathbf{y}) X(\mathbf{y}) = \mathbf{w}_i^\top \mathbf{x} \quad (6)$$

where vectors $\mathbf{x}$ and $\mathbf{w}_i$ contain the pixels of X and elements of W_i, respectively while $(.)^\top$ denotes vector transpose. The binarized feature b_i is obtained by thresholding the filter response s_i at zero

$$b_i = \begin{cases} 1 & s_i > 0, \\ 0 & \text{otherwise.} \end{cases} \quad (7)$$

Given an image patch X of size $l \times l$ pixels, N filters may be applied to X using the filter matrix $W^{N \times l^2}$ to yield N responses which are stacked to constitute vector $\mathbf{s}$. The filter responses are independently binarized next to form an N-bit binary code for each pixel which is then represented as an integer. The codes obtained are finally represented using histogram statistics. Learning BSIF filters entails performing an ICA analysis on the training data. For a detailed discussion on learning BSIF filters one may refer to [38] and [44].

In [38], the BSIF approach is extended to the spatio-temporal domain by considering an image sequence on three orthogonal planes, similar to the LBP-TOP approach [40]. Extension of the BSIF representation to the spatio-temporal domain involves learning three sets of filters, each specific to a certain plane. Once the learned filters are applied to each individual plane, the filter responses are summarized using histograms. Finally, these three histograms are $L1-$ normalised to yield a probability distribution and concatenated to form the final BSIF-TOP descriptor. The multi-scale extension of the approach involves applying filters of different sizes and concatenating corresponding histograms to form the final descriptor. In this work, the filters are applied in six scales. That is, the filters considered are of sizes $\{3 \times 3, 5 \times 5, \ldots, 13 \times 13\}$. The BSIF-TOP dynamic texture representation has been shown to be effective in face spoofing detection in [35].

4) IMAGE QUALITY MEASURES

The fourth set of features used in this work are based on image quality measures. Image quality analysis has been successfully used in different works such as image manipulation detection [45] and steganalysis [46] in the forensic field and more recently for spoofing detection [41]. Employing such measures for face spoofing detection is based on the assumption that capturing an image of a picture displayed on a 2D device can be considered as a type of image manipulation which is detectable using image quality measures. The method for biometrics spoofing detection proposed in [41] uses 25 image quality measures comprised of both reference-based and blind measures. Reference-based measures rely on the availability of an ideal undistorted reference image against which the quality of a test sample is compared. In the problem of spoofing detection such a reference image does not exist. This limitation can, however, be circumvented following the same strategy used in [41]. For this purpose, in [41] the input grey-scale image is first filtered with a low-pass Gaussian kernel. Next, the relative quality of the original and filtered images is assessed using a reference-based quality metric. This approach is based on the premise that

TABLE 1. Image quality measures adopted from [41]. (R denotes reference-based methods while B stands for Blind (no-reference) methods).

#	Type	Name
1	R	Mean Square Error
2	R	Peak Signal to Noise Ratio
3	R	Signal to Noise Ratio
4	R	Structural Content
5	R	Maximum Difference
6	R	Average Difference
7	R	Normalized Absolute Error
8	R	R-Averaged Maximum Difference
9	R	Laplacian Mean Square Error
10	R	Normalized Cross-Correlation
11	R	Mean Angle Similarity
12	R	Mean Angle Magnitude Similarity
13	R	Total Edge Difference
14	R	Total Corner Difference
15	R	Spectral Magnitude Error
16	R	Spectral Phase Error
17	R	Gradient Magnitude Error
18	R	Gradient Phase Error
19	R	Structural Similarity Index
20	R	Visual Information Fidelity
21	R	Reduced Ref. Entropic Difference
22	B	JPEG Quality Index
23	B	High-Low Frequency Index
24	B	Blind Image Quality Index
25	B	Naturalness Image Quality Estimator

the loss of quality incurred by Gaussian filtering differs for real and fake biometric samples. The reference-based quality measures used in [41] comprise error sensitivity measures, structural similarity measures and information theoretic measures.

Unlike reference-based quality measures, blind (no-reference) image quality measures try to assess the visual quality of an image in the absence of a reference sample according to some statistical models. Depending on the images used to train the model and on a priori assumptions made, these methods are coarsely divided into one of three groups as distortion-specific approaches, training-based approaches and natural scene statistic approaches. A list of the 25 image quality measures used in this work is provided in Table 1. For a more detailed description of these measures the reader is referred to [41].

B. OUTLIER DETECTION MECHANISMS

The common notion of anomaly as an outlier from some known class representing normality is referred to as a point anomaly in the literature [6]. The categorisation of methods applicable to detecting point anomalies has been introduced in preceding surveys [6]. These methods are generally classified as statistical, nearest neighbour, classification, clustering, information theoretic or spectral. A more recent and general categorisation [15] considers anomaly detection methods to be of either generative or non-generative nature. While for generative models there is a transparent link between data and models, non-generative models lack a direct relationship to the data. These methods are exemplified by discriminative models which aim to identify the class identity of a pattern by partitioning the observation space. In the current work, both approaches are studied for the task of face spoofing detection. The non-generative models are represented by the one-class SVM classifier while the generative class is exemplified by a one-class sparse representation-based model to detect outliers.

1) ONE-CLASS SVMs

The one-class SVM method of [47], called the Support Vector Data Description (SVDD) approach, is a method for directly obtaining the boundary containing a target set of data samples. In its most simplest form, a hypersphere is estimated which encompass all target samples. The volume of this hypersphere is minimised in order to reduce the chance of accepting outlier samples. In [47], it is shown that the SVDD model can be written in a form comparable to the support vector classifier (SVC) of Vapnik [48]. The SVDD approach also offers the capability of mapping data to a high dimensional feature space as a result of which more flexible descriptions can be obtained. It is known that the SVDD approach provides solutions similar to the hyperplane approach of [49]. The minimisation problem reflecting this approach incorporates both structural and empirical error terms. The structural error corresponds to the radius of the sphere encapsulating target samples while the empirical error term allows the possibility of outliers in the training set. Once an expression for the centre of the hypersphere is derived via the SVDD approach, a new sample can be tested to see whether it is accepted by the description. For this purpose, the distance from the test data sample to the centre of the hypersphere is calculated.

2) SPARSE REPRESENTATION-BASED CLASSIFICATION

In the general concept of linear modelling where a test sample is expressed as a linear combination of available atoms, the sparse representation (SR) method is one of the most representative approaches [50]. SR has been found to be a valuable tool in a variety of different domains including signal and image processing, machine learning, computer vision, etc. One particularly appealing application area of SR is image classification where the sparse representation based classification (SRC) method [51] has been found to be effective in classifying data corrupted by noise, occlusion and some other undesirable perturbations. The SRC assumes that a test sample can be sufficiently well represented as a sparse linear combination of available training samples. In the context of the current work, a descriptor derived from a test sequence is approximated as a sparse linear combination of all training samples. This is represented as

$$\hat{\boldsymbol{\alpha}} = \arg\min_{\boldsymbol{\alpha}} \|\boldsymbol{\alpha}\|_p \quad s.t.\ \|\boldsymbol{y} - \boldsymbol{X}\boldsymbol{\alpha}\|_2^2 \leq \epsilon \tag{8}$$

where $\boldsymbol{X}$ is the set of training samples, $\boldsymbol{y}$ represents the probe sample, $\boldsymbol{\alpha}$ is the sparse coefficients vector and ϵ is a small threshold.

Depending on p, different algorithms have been proposed to solve the above problem. For the $l1$-minimization

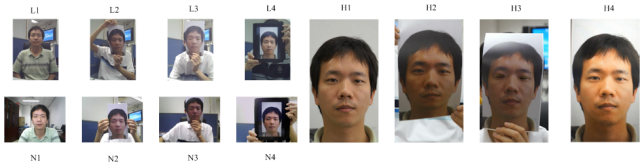

FIGURE 1. Sample data from the CASIA dataset.

problem ($p = 1$) efficient methods are available when the solution is known to be very sparse. Homotopy algorithms [52], [53] are one such class of methods which are shown to run more rapidly than some other alternative solvers.

For multi-class classification, ideally, the non-zero entries in the estimate of the sparse representation solution $\hat{\boldsymbol{\alpha}}$ are expected to be all associated with a single class. In this case, one would easily assign the test sample to the class with non-zero sparse coefficients. However, in practice, noise and modelling errors may cause small non-zero entries associated with multiple classes to exist in the final solution. In this case, one can classify the test sample based on how well it is reproduced via the sparse coefficients associated with the training samples of each individual class. In other words, the reconstruction residual of a test sample using the sparse coefficients of each class is used as a dissimilarity criterion for hypothesis selection:

$$\min_i r_i(y) = \|y - X\delta_i(\hat{\boldsymbol{\alpha}})\|_2^2 \tag{9}$$

In the above equation, $r_i(y)$ estimates the residual for probe y reconstructed as a linear combination of samples of class i and $\delta_i(.)$ is a characteristic function that selects the coefficients associated with the i^{th} class.

Similarly, in a single-class classification problem, the above reconstruction residual can be used as a dissimilarity criterion where typically a threshold is employed for decision making.

IV. EXPERIMENTAL EVALUATION

In this section, first a brief description of the three databases used in the experiments is provided. They comprise samples of printed attacks and video display attacks. Next, a new evaluation protocol is presented. The discussion then focuses on the two-class and one-class systems designed for spoofing detection and the results of the experimental evaluation in different settings.

A. THE CASIA FACE ANTI-SPOOFING DATABASE (CASIA FASD)

The CASIA FASD [29] consists of video sequences of both real and fake access attempts captured using different cameras resulting in videos of different qualities in Fig. 1. Seven test scenarios are defined for the CASIA FASD. These include the warped photo test, cut photo test and video test in addition to the quality (low quality, normal quality and high quality) and overall tests described next.

Warped photo attack: A 1920 × 1080 image for each subject was captured using a Sony NEX-5 camera. This high resolution image was then printed on a high quality paper. In the warped photo attack scenario, in order to imitate real facial motion, the attacker warps the photo.

Cut photo attack: In this scenario, the eye regions of the aforementioned photos are cut out. An attacker then hides behinds the photo and blinks through the holes of the eye region. In addition, blinking is simulated by moving an intact photo behind the cut photos.

Video attack: In this setting, high resolution videos are displayed using an iPad in front of the system camera.

Quality Test: In this test, the three imaging qualities (low quality, normal quality and high quality) are considered for training and test.

The Overall Test: In this test, all the data (regardless of the quality or attack type) is used for a general and overall evaluation.

This database is split into a training set of 20 subjects and a test set containing 30 subjects. For each of the aforementioned seven test scenarios, the data is selected from the corresponding training and test sets for model training and performance evaluation.

Three different attack types based on the media used for spoofing are identified in the CASIA database. These are *Warped Photo*, *Cut Photo* and *Video* data, explained above.

B. THE REPLAY-ATTACK DATABASE

The Replay-Attack database [19] is one of the most commonly used data sets for evaluating spoofing countermeasures. This database consists of 1200 video recordings of both real-access and attack attempts of 50 subjects. The data set includes 200 real-access videos, 200 print attack videos, 400 phone attack videos, and 400 tablet attack videos. For evaluation purposes, the data set is divided into three subsets of training (360 videos), validation (360 videos) and testing (480 videos). The training and the validation subgroups contain 60 real-access videos and 300 attack videos each while the testing subset contains 80 real-access and 400 attack videos. The training set is used to train a classifier while the validation set is typically used to determine a decision threshold. The test set is only used for performance evaluation. The Replay-Attack database includes different attack types based on the media used for spoofing. These include *Printed photo* (where a colour lase printer is used to produce samples on A4 papers), *Digital photo* (where an iPhone or an iPad is used to display a digital image) and *Video* (where an iPhone or an iPad is used to replay a video).

C. THE MSU MOBILE FACE SPOOFING DATABASE (MSU MFSD)

The MSU MFSD database [33] comprises 440 videos of photo and video attack attempts of 55 individuals in Fig. 2. A laptop camera as well as a front-facing camera in the Google Nexus 5 Android phone, referred to as the Android camera, are used in creating the database. Using the laptop

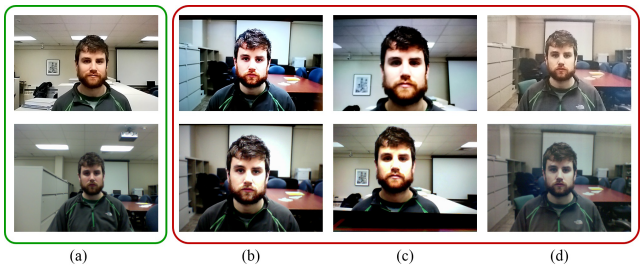

FIGURE 2. Sample data from the MSU dataset.(a) Genuine faces; (b)-(d) Spoof faces.

camera, videos are captured with a resolution of 640 × 480 pixels. For the Android camera, videos are captured with a resolution of 720 × 480 pixels. The average frame rate is 30 fps and the average duration of videos is 12 seconds. In order to simulate a real access operation, the subject faces the camera and a genuine access video is recorded using both the Android and laptop cameras. The average distance between the face and the camera is 50cm. In order to gather spoofing attempt data either a video replay or a printed photograph is used. In the case of video replay, a Canon 550D SLR camera is used to capture a HD video of resolution 1920 × 1088 pixels, which is then replayed on an iPad Air to generate the HD replay attack video. An iPhone 5S is also used to capture another HD video of resolution 1920 × 1080 pixels that is replayed on the iPhone 5S screen to generate the mobile video replay attack. The average stand-off distance for recording the HD video replay attack is 20cm, whereas the average stand-off distance for the mobile video replay attack is 10cm. In the case of a spoofing attack using a printed photo, a Canon 550D camera is used to capture a HD picture of resolution 5184 × 3456 pixels which is then printed on an A3 paper using an HP Colour Laserjet printer (1200 × 600 dpi) to generate a printed photo for attack. The average stand-off distance in this case is 40cm. On the MSU database, different attack types are printed photo attack (where a colour HP laserjet printer is used to produce spoofing data on A3 papers) and video attack. However, as there are a larger number of video attacks compared to other data sets, the video data in this database are further divided into *Mobile phone video* (where an iPhone is used for replaying video) and *High resolution video* (where an iPad is used for video replay).

D. THE PROPOSED EVALUATION PROTOCOLS

The premise behind the current face spoofing detection research is that antispoofing technology can be developed by collecting spoofing attack data for various types of forgery. With a sufficient volume of spoofing data available, antispoofing solutions can then be developed by formulating the detection problem as a conventional two-class discrimination task, i.e. normal access versus spoofing attack, and adopt conventional training procedures to design a spoofing detector. The implicit assumption behind this philosophy is that the responsibility for bringing a spoofing detector technology to the market rests with an independent third party industry rather than with the biometric technology provider. In such a scenario, which we shall refer to as *conventional scenario*, collecting both types of samples, i.e. normal access and spoofing attack samples, is a necessary prerequisite and considered as the normal practice. The acquisition of both categories of accesses is equally difficult and for each subject participating in data collection both types can be acquired as part of the same data collection campaign.

The existing evaluation protocols in the literature [19], [29], [33] reflect this point of view. The protocols have been designed to measure the spoofing detection system performance in two basic conditions. The stable conditions scenario involves training and testing the detection systems on data within a single database where all normal and spoofing attack accesses are acquired using the same sensor set up and deploying an identical implementation of spoofing attempts. The protocol involves splitting all the data available into training and test data and guaging the true and false positive detection rates accordingly. This category of evaluation protocols is known as *intra-database* testing. The second category of tests measures the generalisation capability of the spoofing detection solutions developed using the data of one database for the system design on other databases. This is referred to as the *inter-database* testing protocol.

The above conventional evaluation protocols fail to address an extremely pertinent question. How does a spoofing detection system cope with a new type of spoofing attack. Individuals aiming to deceive biometric systems are very inventive and create new forms of attack which cannot necessarily be envisaged beforehand. In the absence of training data the spoofing attack detectors designed using the conventional two-class training approach are likely to be ineffective. To measure this capacity to detect novel forms of attack, a new protocol is required. To this effect we propose a new family of protocols, named *innovative attack evaluation protocol* designed to provide this crucial performance information. In essence, as the focus is on assessing the generalisation capability of various systems subject to different types of attacks, in each scenario, one attack type is excluded from the training set and presented to the system only during the test phase. The proposed experimental protocols ensure that the spoofing detection systems do not see any spoofing accesses of a new type during training. Thus for intra-database performance evaluation the systems use training data associated with two spoofing accesses plus corresponding normal access data and are then tested on the third type of attack. This is repeated in rotation for the other two types of attack and for the other two databases as well. In contrast, to evaluate the generalisation capacity of the spoofing detection systems in the inter-database testing scenario, the two-class systems are trained using normal and spoofing attack data of two databases and test on the third database. The spoofing attack data of the type tested on the third database is always excluded from training the two-class systems.

The data usage corresponding to the evaluation protocols is summarised in Table 2. In the table the symbol "+"

TABLE 2. Two-class classifier system evaluation.(+:training, ∗:test) Note: for one-class classifier system evaluation only the marked normal class data is use for training.

			Database											
			Casia				Replay-Attack				MSU			
Protocol Scenario			normal	Warped	Cut	Video	Normal	printed	digital	video	normal	printed	Hi-res. Video	Mobile video
Intra	Casia	Warped	+*	*	+	+								
		Cut Photo	+*	+	*	+								
		Video	+*	+	+	*								
	Replay -Attack	Printed					+*	*	+	+				
		Digital					+*	+	*	+				
		Video					+*	+	+	*				
	MSU	Printed									+*	*	+	+
		Hi-res. Video									+*	+	*	+
		mobile video									+*	+	+	*
Inter	Casia	Warped	*	*			+		+	+	+		+	+
		Cut Photo	*		*		+	+	+	+	+	+	+	
		Video	*			*	+	+	+		+	+		
	Replay -Attack	Printed	+		+	+	*	*			+		+	+
		Digital	+	+	+	+	*		*		+	+	+	
		Video	+	+	+		*			*	+	+		
	MSU	Printed	+		+	+	+		+	+	*	*		
		Hi-res. Video	+	+	+		+	+	+		*		*	
		mobile video	+	+	+		+	+	+		*			*

indicates the sections of the three databases used for training, and "*" denotes the sections used for testing. For instance, in the Protocol Scenario designated as Inter-Replay Attack-Video, the two class systems are trained on normal access data in the Casia and MSU datasets, and the Warped and Cut spoofing attack data from the Casia database as well as the Printed spoofing attack from the MSU database. None of the video related spoofing attacks in Casia and MSU are used for training. The testing is performed on the normal and video spoofing attack access of the Replay database.

As already argued, in our opinion, the above conventional design scenario is not completely realistic for several reasons:

1) Biometric technology providers are commercially motivated and technologically best qualified to provide both face recognition and antispoofing technologies, potentially as part of a single integrated system. Under this assumption it is relatively easy to collect biometric access data and extract from it measurements relevant for spoofing attack detection.
2) Normal access data can be collected with relative ease whereas spoofing attack data requires simulating the various forms of attack which is demanding in terms of manpower resources.

Accordingly, a more realistic scenario promulgated in this paper is to formulate the spoofing detection problem as one of anomaly detection where spoofing is viewed as an outlier of the learnt distribution of a spoofing test statistic which is based on some features extracted from video access information collected for genuine users. The assumption is that any spoofing attack will be distinguishable from normal accesses based on this test statistic. In complience with this new philosophy to spoofing attack detector design, only normal access data is required for training, and the detection engine is based on the one-class classification approach. The aim of this paper is to investigate the performance of spoofing detectors based on this one-class classifier technique and their ability to detect spoofing attacks regardless of the type. As the design of one-class systems is based on much less training data (normal accesses only) than that used for their two-class counterparts, their performance measured using conventional evaluation protocols can be expected to be inferior. However, their ability to detect unknown spoofing attacks could be better than that of two-class systems, as in essence one-class systems aim to encapsulate the normal access data and any deviations from the norm, including novel departures should be detectable.

The experimental protocol for testing one-class spoofing detection systems is very simple. They are trained using normal access training data only. In the intra-database scenario the decision making systems are trained using the normal access training data available in one of the spoofing databases. Their performance is measured using the normal access test data, and spoofing attack test data of the same database. The detection rate for each type of spoofing attack is measured separately. This experiment is repeated for each of the three databases. In the inter-database evaluation we train one-class classifier systems using normal access training data from two databases, and test on the normal access and spoofing access test data of the third database. The performance for each type of attack is recorded separately. The process is repeated for each spoofing attack type and then for each spoofing database.

It should be reiterated that the one-class systems examined in this work do not use negative training data (spoofing attacks). Only the two class systems employ such data.

E. TWO-CLASS SYSTEMS

The two-class systems evaluated are constructed using a two-class SVM with a radial basis function, the two-class LDA classifier and the sparse representation-based two-class classifier. For the LDA-based systems, a PCA transformation is initially applied to the data in order to make the within-class scatter matrix invertible. For the sparse-representation-based classifier, the Homotopy algorithm is used [52]. In these systems, both real accesses as well as spoofing attacks are used for training a classifier. In particular, the following systems are evaluated:

- SVM2+BSIF-TOP: The two-class SVM classifier with a Gaussian kernel trained using the multi-scale BSIF-TOP features.
- SVM2+LPQ-TOP: The two-class SVM classifier with a Gaussian kernel trained using the multi-scale LPQ-TOP features.
- SVM2+LBP-TOP: The two-class SVM classifier with a Gaussian kernel trained using the multi-scale LBP-TOP features.
- SVM2+IMQ: The two-class SVM classifier with a Gaussian kernel trained using the image quality features.
- LDA2+BSIF-TOP: The two-class LDA classifier trained using the multi-scale BSIF-TOP representations.
- LDA2+LPQ-TOP: The two-class LDA classifier trained using the multi-scale LPQ-TOP representations.
- LDA2+LBP-TOP: The two-class LDA classifier trained using the multi-scale LBP-TOP representations.
- LDA2+IMQ: The two-class LDA classifier trained using image quality features.
- SRC2+BSIF-TOP: The two-class sparse representation-based classifier trained using the multi-scale BSIF-TOP representations.
- SRC2+LPQ-TOP: The two-class sparse representation-based classifier trained using the multi-scale LPQ-TOP representations.
- SRC2+LBP-TOP: The two-class sparse representation-based classifier trained using the multi-scale LBP-TOP representations.
- SRC2+IMQ: The two-class sparse representation-based classifier trained using image quality features.

F. ONE-CLASS SYSTEMS

As noted earlier, the systems evaluated for anomaly detection in this work are based on the one-class SVM [47] and the sparse representation-based classifier [51]. In particular, the following systems are evaluated for one-class anomaly detection:

- SVM1+BSIF-TOP: The one-class SVM with a Gaussian kernel trained using the multi-scale BSIF-TOP features.
- SVM1+LPQ-TOP: The one-class SVM with a Gaussian kernel trained using the multi-scale LPQ-TOP features.
- SVM1+LBP-TOP: The one-class SVM with a Gaussian kernel trained using the multi-scale LBP-TOP features.
- SVM1+IMQ: The one-class SVM with a Gaussian kernel trained using the image quality features.
- SRC1+BSIF-TOP: The sparse representation-based one-class classifier trained using the multi-scale BSIF-TOP features.
- SRC1+LPQ-TOP: The sparse representation-based one-class classifier trained using the multi-scale LPQ-TOP features.
- SRC1+LBP-TOP: The sparse representation-based one-class classifier trained using the multi-scale LBP-TOP features.
- SRC1+IMQ: The sparse representation-based one-class classifier trained using the image quality features.

G. IMPLEMENTATION DETAILS

A few comments regarding the implementation of the above systems are in order. The colour frames of videos of all databases are converted to gray scale images before performing any further processing. The representations used are extracted in two different settings: once from the whole image and next from the face region only. In order to extract the face regions, the Viola-Jones face detection algorithm is run on each sequence starting from the first frame. The location of the first detected face in a sequence is then used for all frames of the video under consideration.

Dynamic texture features are extracted in six different scales for each of the LBP-TOP, LPQ-TOP and the BSIF-TOP representations. For the BSIF-TOP representation, in order to make the results reproducible, we have used the filters provided in [44] for each of the three orthogonal planes. Regarding the image quality measures, we have used the original source code provided by Galbally *et al.* [41]. As image quality features are extracted from a single frame, in order to make use of the whole image sequence, we have extracted them from 20 evenly spaced frames of an image sequence. The resulting feature vectors are then averaged to produce the final feature vector for the whole sequence.

The two-class SVM and the one-class SVDD approach are implemented using the Libsvm library [54]. The kernel functions used in these methods are the radial basis function ($e^{\|x-y\|^2/2\sigma^2}$). The σ parameter in the Gaussian kernel function is determined by cross validation on the training set and a simple grid search. Regarding the LDA classifier, the dimensionality of the PCA transformation is determined by retaining 98% of the variation present in the data.

In the operational mode of a face spoofing detection system, one should set a threshold against which a test sample is compared. Setting such a threshold is typically performed using an independent set of evaluation data. As only the Replay-Attack database contains such an evaluation set, we report the performance of all systems using the area under the ROC curve measure for a fair and reproducible comparison. Such an approach removes the complications associated with setting a specific threshold and provides a good measure of *average system performance* irrespective of a certain decision threshold.

TABLE 3. Area under the ROC (AUC) (%) for different systems in the intra-database & cross type setting on the CASIA database.

System	Warped photo	Cut photo	Video
SVM2+IMQ(w)	84.58	89.60	94.82
SVM2+IMQ(f)	75.24	60.21	70.33
SVM2+BSIF(w)	93.21	70.84	79.43
SVM2+BSIF(f)	89.25	48.10	69.36
SVM2+LPQ(w)	84.14	64.83	74.74
SVM2+LPQ(f)	86.55	55.56	77.65
SVM2+LBP(w)	84.91	63.99	77.31
SVM2+LBP(f)	79.57	39.55	67.10
SRC2+IMQ(w)	83.85	79.86	90.41
SRC2+IMQ(f)	81.44	68.96	71.72
SRC2+BSIF(w)	90.15	77.06	90.59
SRC2+BSIF(f)	81.78	55.72	71.54
SRC2+LPQ(w)	84.36	69.81	76.79
SRC2+LPQ(f)	77.91	61.88	72.51
SRC2+LBP(w)	83.48	73.04	87.05
SRC2+LBP(f)	74.94	51.74	67.68
LDA2+IMQ(w)	66.83	65.80	93.20
LDA2+IMQ(f)	69.40	61.72	56.96
LDA2+BSIF(w)	65.46	53.65	57.26
LDA2+BSIF(f)	38.16	44.93	70.25
LDA2+LPQ(w)	70.53	56.17	65.25
LDA2+LPQ(f)	18.85	46.14	72.21
LDA2+LBP(w)	66.07	57.32	65.58
LDA2+LBP(f)	16.56	37.83	61.37
SVM1+IMQ(w)	76.93	72.00	86.46
SVM1+IMQ(f)	74.80	61.95	68.89
SVM1+BSIF(w)	81.05	50.78	80.38
SVM1+BSIF(f)	95.90	60.73	70.74
SVM1+LPQ(w)	85.26	54.33	78.84
SVM1+LPQ(f)	96.49	59.58	74.73
SVM1+LBP(w)	84.89	51.32	91.58
SVM1+LBP(f)	95.23	66.40	72.73
SRC1+IMQ(w)	80.59	79.94	92.75
SRC1+IMQ(f)	66.85	67.10	65.43
SRC1+BSIF(w)	86.35	64.10	90.30
SRC1+BSIF(f)	94.85	76.36	78.88
SRC1+LPQ(w)	86.86	65.01	89.85
SRC1+LPQ(f)	95.32	67.96	80.59
SRC1+LBP(w)	83.99	64.51	73.46
SRC1+LBP(f)	92.79	62.56	80.59

TABLE 4. Area under the ROC (AUC) (%) for different systems in the intra-database & cross type setting on the Replay-Attack database.

System	Printed photo	Digital photo	Video
SVM2+IMQ(w)	73.56	89.15	98.23
SVM2+IMQ(f)	49.31	83.67	97.96
SVM2+BSIF(w)	96.98	99.05	99.06
SVM2+BSIF(f)	90.10	97.48	97.25
SVM2+LPQ(w)	95.97	97.84	98.90
SVM2+LPQ(f)	91.56	95.80	95.32
SVM2+LBP(w)	97.37	98.57	99.06
SVM2+LBP(f)	87.37	96.57	93.86
SRC2+IMQ(w)	78.34	93.46	94.66
SRC2+IMQ(f)	44.44	85.34	95.45
SRC2+BSIF(w)	98.75	99.38	99.38
SRC2+BSIF(f)	91.66	98.15	98.37
SRC2+LPQ(w)	97.41	99.09	99.32
SRC2+LPQ(f)	94.75	96.08	94.85
SRC2+LBP(w)	98.22	99.09	99.38
SRC2+LBP(f)	90.33	98.34	95.92
LDA2+IMQ(w)	37.50	94.06	92.21
LDA2+IMQ(f)	33.41	92.26	96.09
LDA2+BSIF(w)	93.72	85.38	76.13
LDA2+BSIF(f)	87.61	82.97	73.78
LDA2+LPQ(w)	84.27	68.33	57.72
LDA2+LPQ(f)	78.19	63.23	51.20
LDA2+LBP(w)	97.72	94.16	85.27
LDA2+LBP(f)	84.69	71.81	59.88
SVM1+IMQ(w)	73.05	96.86	99.34
SVM1+IMQ(f)	53.23	90.82	98.24
SVM1+BSIF(w)	91.19	99.04	98.95
SVM1+BSIF(f)	73.66	88.14	84.03
SVM1+LPQ(w)	89.47	94.74	95.68
SVM1+LPQ(f)	73.89	85.70	83.04
SVM1+LBP(w)	97.98	99.06	98.77
SVM1+LBP(f)	70.62	84.68	77.74
SRC1+IMQ(w)	77.05	96.38	98.16
SRC1+IMQ(f)	55.38	87.62	96.58
SRC1+BSIF(w)	98.41	99.29	99.25
SRC1+BSIF(f)	88.17	94.88	94.02
SRC1+LPQ(w)	97.89	98.69	99.35
SRC1+LPQ(f)	73.70	86.42	93.02
SRC1+LBP(w)	98.39	99.06	99.33
SRC1+LBP(f)	84.66	91.31	95.52

H. EVALUATION RESULTS FOR THE INTRA-DATABASE & CROSS-TYPE SETTING

The one-class and two-class systems introduced earlier are evaluated on the three databases of CASIA, Replay-Attack and MSU in the intra-database & cross-type setting for a total of 9 different evaluations. As noted earlier, in this experiment, since the training and test data come from the same database, the focus here is on the effect of the type of spoofing media on the system performance. In this setting, two types of tests are performed: whole image test and the face region test. In the whole image test, each frame of a video sequence is used as a whole for extracting features whereas in the face region test, only the face region in each frame is used to construct the feature vector. The results of this experiment are reported in Tables 3, 4 and 5 for the CASIA, Replay-Attack and MSU datasets, respectively. The performance of different systems is summarised using mean and standard deviation of area under the ROC curves in Table 6. In the tables, ''w'' denotes a whole image test whereas ''f'' stands for a face region test. Examining Tables 3, 4 and 5 reveals that in general, in the intra-database & cross-type setting, the systems operating on the whole image perform better than the systems operating on the face region only. This is intuitive as in the intra-database setting the background remains almost the same and can be exploited to improve performance.

As can be seen from Table 6, the best performing two-class system in terms of average performance is SRC2+BSIF(w) with an average AUC of 92.51%. The best performing method among the one-class systems is SRC1+BSIF(w) with an average AUC of 90.77%. The average performance of the best performing one-class method in this case is only less than 2% worse than the best performing two-class approach.

If one restricts the attention only to the systems operating on the face region, the best performing two-class approach is SVM2+LPQ(f) with an average AUC of 81.53% whereas the best performing one-class method in this case is SRC1+IMQ(f) with an average performance of 87.84% which is more than 6% better than the two-class alternative. These observations clearly illustrate the potential of the one-class anomaly detection approach in the challenging conditions of cross-type evaluations.

I. EVALUATION RESULTS FOR THE INTER-DATABASE & CROSS-TYPE SETTING

In this section, the one-class and two-class systems introduced earlier are evaluated on the three databases of CASIA, Replay-Attack and MSU in the *inter-database & cross-type* setting for a total of 9 different evaluation schemes. The results of this experiment are reported in Table 7 and also summarised as mean and standard deviations of area under the ROC curves in Table 8. As can be seen from Table 8, in this experiment, similar to the intra-database setting, using the background improves performance when using image quality features. However, in contrast to the intra-database setting, using the background deteriorates the performance when using dynamic texture descriptors. This is expected as in the inter-database setting, there exist different backgrounds between the train and test samples which adversely affect the performance.

TABLE 5. Area under the ROC (AUC) (%) for different systems in the intra-database & cross type setting on the MSU database.

System	Printed photo	Digital photo	Video
SVM2+IMQ(w)	22.94	39.97	88.16
SVM2+IMQ(f)	47.19	36.94	69.22
SVM2+BSIF(w)	85.09	95.66	90.94
SVM2+BSIF(f)	61.06	92.56	54.00
SVM2+LPQ(w)	86.88	95.81	84.19
SVM2+LPQ(f)	74.41	93.97	62.97
SVM2+LBP(w)	91.03	96.22	91.16
SVM2+LBP(f)	58.25	90.81	37.91
SRC2+IMQ(w)	40.75	67.72	46.75
SRC2+IMQ(f)	47.12	55.94	83.38
SRC2+BSIF(w)	85.19	97.50	94.63
SRC2+BSIF(f)	70.12	93.00	63.62
SRC2+LPQ(w)	88.94	97.25	89.88
SRC2+LPQ(f)	72.06	92.63	57.94
SRC2+LBP(w)	91.94	97.50	93.06
SRC2+LBP(f)	47.62	86.38	53.06
LDA2+IMQ(w)	68.13	40.44	63.81
LDA2+IMQ(f)	53.19	52.00	84.63
LDA2+BSIF(w)	76.25	92.25	95.56
LDA2+BSIF(f)	58.25	78.81	48.88
LDA2+LPQ(w)	81.50	94.13	88.38
LDA2+LPQ(f)	75.69	84.19	43.87
LDA2+LBP(w)	81.31	74.50	74.38
LDA2+LBP(f)	30.75	67.56	27.25
SVM1+IMQ(w)	86.13	74.50	72.50
SVM1+IMQ(f)	63.94	63.00	76.38
SVM1+BSIF(w)	88.31	95.06	94.06
SVM1+BSIF(f)	64.81	87.44	74.69
SVM1+LPQ(w)	89.44	94.13	90.56
SVM1+LPQ(f)	74.94	91.06	76.19
SVM1+LBP(w)	93.88	96.06	96.19
SVM1+LBP(f)	60.69	63.31	68.38
SRC1+IMQ(w)	86.56	91.00	88.19
SRC1+IMQ(f)	68.56	69.94	69.25
SRC1+BSIF(w)	86.81	96.37	96.12
SRC1+BSIF(f)	83.69	88.25	83.38
SRC1+LPQ(w)	73.94	93.50	92.81
SRC1+LPQ(f)	63.31	84.25	84.38
SRC1+LBP(w)	93.25	96.63	96.81
SRC1+LBP(f)	56.37	59.31	65.81

TABLE 6. Mean and standard deviations of Area under the ROC (AUC) (%) measures for different systems in the intra-database & cross type setting obtained on three data sets.

System	Mean	Std.
SVM2+IMQ(f)	65.56	19.22
SVM2+IMQ(w)	75.66	26.33
SVM2+BSIF(f)	77.68	19.57
SVM2+BSIF(w)	90.02	9.73
SVM2+LPQ(f)	81.53	14.82
SVM2+LPQ(w)	87.03	11.59
SVM2+LBP(f)	72.33	22.79
SVM2+LBP(w)	88.84	11.73
SRC2+IMQ(f)	70.42	17.92
SRC2+IMQ(w)	75.08	19.69
SRC2+BSIF(f)	80.44	15.81
SRC2+BSIF(w)	92.51	7.61
SRC2+LPQ(f)	80.06	14.97
SRC2+LPQ(w)	89.20	10.52
SRC2+LBP(f)	74.00	19.88
SRC2+LBP(w)	91.41	8.86
LDA2+IMQ(f)	66.62	20.83
LDA2+IMQ(w)	69.10	21.22
LDA2+BSIF(f)	69.71	23.37
LDA2+BSIF(w)	77.29	15.79
LDA2+LPQ(f)	59.28	21.03
LDA2+LPQ(w)	75.15	14.64
LDA2+LBP(f)	50.85	23.35
LDA2+LBP(w)	77.36	13.50
SVM1+IMQ(f)	72.36	14.48
SVM1+IMQ(w)	81.97	10.66
SVM1+BSIF(f)	77.79	11.74
SVM1+BSIF(w)	86.53	15.04
SVM1+LPQ(f)	79.51	10.93
SVM1+LPQ(w)	85.82	12.91
SVM1+LBP(f)	73.30	11.00
SVM1+LBP(w)	89.97	15.15
SRC1+IMQ(f)	71.85	12.45
SRC1+IMQ(w)	87.84	7.47
SRC1+BSIF(f)	86.94	6.88
SRC1+BSIF(w)	90.77	11.23
SRC1+LPQ(f)	80.99	10.82
SRC1+LPQ(w)	88.65	11.83
SRC1+LBP(f)	76.54	15.56
SRC1+LBP(w)	89.49	12.73

The best performing one-class system in this case is SVM1+IMQ(w) (the one-class SVM operating on the image quality features extracted from the whole image) with an average performance of 70.23%. In comparison, the best performing two-class system is LDA2+IMQ(w) (the two-class LDA operating on image quality features extracted from the whole image) with an average performance of 69.38%. It can be concluded that the image quality features in the challenging conditions of inter-database and cross-type evaluations are more robust as compared with the dynamic texture descriptors. More importantly, the one-class anomaly-based systems are not inferior compared with two-class alternatives.

J. DISCUSSION

A number of comments regarding the key findings are in order.

- In terms of representations, we have examined three different dynamic texture descriptors and quality-based features. In the intra-database setting, the BSIF-TOP representation achieved the best performance among other representations both in the one-class and two-class systems. However, in the more challenging conditions of inter-database evaluation, the image quality features proved to be more robust than the dynamic texture descriptors.
- Regarding the use of background information, in the intra-database evaluation, the performance of both one-class and two-class systems using either one of the four different representations improved. However, in the inter-database setting, the use of background deteriorated the performances of systems operating on dynamic texture descriptors. The image quality features, however, in the inter-database setting could benefit from background information.
- The anomaly detection-based systems in both intra- and inter-database settings are not inferior as compared with the two-class alternatives. As we have only examined outlier detection systems based on a one-class SVM and a sparse representation model, it would be beneficial to study other anomaly detection approaches.
- In this paper we examined the effectiveness of one-class SVM trained on positive samples only. While it is quite unrealistic to assume that spoofing attacks samples are abundantly available, in a real scenario there will be attempts at spoofing and samples will become progressively available over time. It would be interesting to investigate the merit of using these sparse negative samples to refine the one-class SVM solution as suggested by Tax and Duin in [55].
- The effects of additional normal data and subject specific training are to be investigated.
- The anomaly detection approach considered in this work may be extended to the domain anomaly concept proposed in [15]. In this case, one may examine the quality of the data before making a decision. In this respect, if the quality of the test data is very different from those of training, the decision may not be very reliable and system disables the anomaly detection mechanism.

TABLE 7. Area under the ROC (AUC) (%) for different systems in the inter-database & cross type setting.

System	Video (C.)	Cut Photo (C.)	Warped Photo (C.)	Video (R.)	Digital Photo (R.)	Printed Photo (R.)	Printed Photo (M.)	High Resolution Video (M.)	Mobile Phone Video (M.)
SVM2+IMQ(f)	43.72	49.47	46.94	17.16	70.53	56.07	43.81	61.34	87.56
SVM2+IMQ(w)	63.86	66.35	65.72	92.84	73.39	21.72	71.12	75.22	64.72
SVM2+BSIF(f)	50.91	57.22	77.58	21.17	57.94	59.97	33.34	67.94	59.66
SVM2+BSIF(w)	51.42	53.79	50.98	75.22	62.77	51.14	69.44	52.06	63.69
SVM2+LPQ(f)	62.27	57.43	83.67	30.46	60.06	56.91	39.69	65.94	78.59
SVM2+LPQ(w)	40.61	55.54	60.96	65.97	41.72	36.80	60.50	28.25	40.91
SVM2+LBP(f)	34.46	36.19	77.60	15.43	16.02	25.16	30.00	41.66	55.53
SVM2+LBP(w)	37.58	57.62	53.81	93.80	53.73	41.97	77.22	50.12	66.47
SRC2+IMQ(f)	67.11	65.49	51.81	36.23	60.61	52.58	56.50	61.66	80.87
SRC2+IMQ(w)	60.08	62.89	70.25	97.49	90.73	30.66	79.50	66.56	63.81
SRC2+BSIF(f)	53.74	55.23	76.17	29.98	48.69	40.67	36.19	58.13	55.50
SRC2+BSIF(w)	57.11	60.90	59.99	81.64	64.04	45.48	67.31	60.25	78.19
SRC2+LPQ(f)	55.48	55.05	65.72	31.65	38.46	38.34	35.94	42.25	65.50
SRC2+LPQ(w)	41.01	57.51	56.96	67.16	46.98	51.36	41.69	50.44	73.25
SRC2+LBP(f)	56.74	37.41	69.84	34.42	35.03	35.28	34.38	36.44	42.44
SRC2+LBP(w)	46.02	56.96	63.73	69.33	46.18	59.67	70.44	68.12	80.19
LDA2+IMQ(f)	41.69	39.68	44.85	41.34	73.76	38.02	26.06	29.00	61.13
LDA2+IMQ(w)	81.20	64.00	61.37	82.39	92.51	45.77	68.81	47.00	81.44
LDA2+BSIF(f)	35.53	31.32	25.01	88.56	43.71	60.55	25.06	23.50	53.88
LDA2+BSIF(w)	31.60	60.15	72.56	64.91	54.43	45.53	44.06	30.50	36.50
LDA2+LPQ(f)	30.88	47.00	25.28	89.41	45.19	65.25	22.56	25.38	53.63
LDA2+LPQ(w)	16.11	57.48	93.60	65.55	35.56	27.19	23.50	11.56	17.19
LDA2+LBP(f)	37.51	26.56	5.27	54.46	18.10	35.02	27.62	24.75	22.13
LDA2+LBP(w)	28.89	55.56	36.31	82.15	25.09	28.89	74.50	47.81	44.00
SVM1+IMQ(f)	63.26	59.43	66.81	84.48	67.57	70.30	53.94	84.75	76.56
SVM1+IMQ(w)	88.41	75.14	75.23	88.21	71.20	56.41	56.62	71.12	49.75
SVM1+BSIF(f)	67.59	51.01	96.33	46.54	63.24	38.88	62.06	80.56	64.06
SVM1+BSIF(w)	64.65	43.16	49.67	71.55	69.70	18.06	67.06	82.88	88.25
SVM1+LPQ(f)	70.64	54.01	96.64	47.09	64.28	44.48	72.38	81.81	74.12
SVM1+LPQ(w)	52.47	40.40	48.17	65.30	53.32	15.09	46.37	56.63	73.19
SVM1+LBP(f)	67.83	55.67	95.57	50.23	61.93	41.05	56.19	47.25	65.00
SVM1+LBP(w)	43.69	48.78	49.53	63.85	71.88	21.97	68.81	62.87	87.25
SRC1+IMQ(f)	62.06	60.75	60.86	18.91	19.18	50.75	59.75	69.56	67.37
SRC1+IMQ(w)	67.68	55.36	63.02	56.38	44.25	27.63	72.69	74.06	70.13
SRC1+BSIF(f)	54.28	69.48	96.57	43.31	51.62	32.20	71.63	76.63	69.63
SRC1+BSIF(w)	70.68	54.07	78.75	78.25	61.21	38.48	45.25	40.75	78.50
SRC1+LPQ(f)	65.93	56.51	97.00	52.88	45.58	28.05	43.87	64.00	76.81
SRC1+LPQ(w)	83.14	59.01	78.68	87.02	59.80	26.78	33.88	43.94	58.12
SRC1+LBP(f)	61.60	56.10	97.17	69.56	65.46	44.17	40.56	35.44	80.06
SRC1+LBP(w)	46.41	55.99	73.33	95.01	81.85	81.58	57.44	53.13	81.13

TABLE 8. Mean and standard deviations of the Area under the ROC (AUC) (%) for different systems in the inter-database & cross type setting.

System	Mean	Std.
SVM2+IMQ(f)	52.95	19.63
SVM2+IMQ(w)	66.10	18.88
SVM2+BSIF(f)	53.97	17.16
SVM2+BSIF(w)	58.94	9.13
SVM2+LPQ(f)	59.44	16.751
SVM2+LPQ(w)	47.91	13.05
SVM2+LBP(f)	36.89	19.73
SVM2+LBP(w)	59.14	17.61
SRC2+IMQ(f)	59.20	12.31
SRC2+IMQ(w)	69.10	19.41
SRC2+BSIF(f)	50.47	13.69
SRC2+BSIF(w)	63.87	10.91
SRC2+LPQ(f)	47.59	13.01
SRC2+LPQ(w)	54.04	10.92
SRC2+LBP(f)	42.44	12.51
SRC2+LBP(w)	62.29	11.35
LDA2+IMQ(f)	43.94	14.97
LDA2+IMQ(w)	69.38	16.35
LDA2+BSIF(f)	43.01	21.56
LDA2+BSIF(w)	48.91	14.99
LDA2+LPQ(f)	44.95	22.18
LDA2+LPQ(w)	38.63	27.77
LDA2+LBP(f)	27.93	13.70
LDA2+LBP(w)	47.02	20.38
SVM1+IMQ(f)	69.67	10.61
SVM1+IMQ(w)	70.23	13.69
SVM1+BSIF(f)	63.36	17.45
SVM1+BSIF(w)	61.66	21.63
SVM1+LPQ(f)	67.27	16.84
SVM1+LPQ(w)	50.10	16.43
SVM1+LBP(f)	60.08	15.83
SVM1+LBP(w)	57.62	18.97
SRC1+IMQ(f)	52.13	19.47
SRC1+IMQ(w)	59.02	15.21
SRC1+BSIF(f)	62.81	19.41
SRC1+BSIF(w)	60.66	16.67
SRC1+LPQ(f)	58.95	20.13
SRC1+LPQ(w)	58.93	21.36

- As the current work aimed to provide a baseline for the proposed anomaly-based formulation of the face spoofing detection problem, standard representations including dynamic texture descriptors and image quality features which are well established feature descriptors were used in the evaluations. Nevertheless, more recent representations such as those in [56]–[59] might be investigated for improved performance.
- It can be concluded that neither the two-class systems nor the one-class approaches perform well enough and more research should be conducted to enhance current systems.

V. CONCLUSIONS

The paper addressed the face spoofing detection problem. Motivated by different observations such as the diversity of spoofing attacks, any new means of spoofing attackers may invent (previously unseen by the system), the problem of imaging sensor inter-operability and other environmental factors in addition to the small sample size, a number of propositions are put forth. First, based on the aforementioned factors, a new evaluation protocol to study the effect of occurrence of unseen attack types was proposed. In the new evaluation scheme, all samples from the same type as that of a test sample were excluded from the training set. Both inter-database and intra-database experiments were then conducted using the proposed evaluation scheme to account for variability in imaging conditions.

In terms of conceptual innovations, a new and more realistic formulation of the spoofing detection problem based on the anomaly detection concept was proposed. The new formulation only required positive samples for system training. Such a one-class formulation eliminated the need for the availability of negative training samples, which, from the generalisation point of view, should be representative of all possible spoofing types. Finally, a thorough evaluation and comparison of 20 different one-class and two-class

systems based on common spatio-temporal as well as image quality features was performed. It was demonstrated that the anomaly-based formulation is not inferior as compared with the conventional two-class approach. Other imaging conditions such using background information were also investigated. Finally, it was concluded that the performance of both formulations (one-class and two-class) was not adequate and more research was required to enhance the detection rates.

ACKNOWLEDGMENT

W. Christmas, deceased, was with the Centre for Vision, Speech and Signal Processing, University of Surrey, Guildford GU2 7XH, U.K.

REFERENCES

[1] T. de Freitas Pereira *et al.*, "Face liveness detection using dynamic texture," *EURASIP J. Image Video Process.*, vol. 2014, no. 1, p. 2, Jan. 2014.

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

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

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

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

[6] V. Chandola, A. Banerjee, and V. Kumar, "Anomaly detection: A survey," *ACM Comput. Surv.*, vol. 41, no. 3, p. 15, Jul. 2009.

[7] F. Edgeworth, "On discordant observations," *Philos. Mag.*, vol. 23, no. 5, pp. 364–375, 1887.

[8] T. Lane and C. E. Brodley, "Temporal sequence learning and data reduction for anomaly detection," *ACM Trans. Inf. Syst. Secur.*, vol. 2, no. 3, pp. 295–331, 1999.

[9] H. Dutta, C. Giannella, K. D. Borne, and H. Kargupta, "Distributed top-k outlier detection from astronomy catalogs using the DEMAC system," in *Proc. SDM SIAM*, 2007, pp. 473–478.

[10] M. Augusteijn and B. Folkert, "Neural network classification and novelty detection," *Int. J. Remote Sens.*, vol. 23, no. 14, pp. 2891–2902, 2002.

[11] Y. L. Cun *et al.*, "Handwritten digit recognition with a back-propagation network," in *Proc. Adv. Neural Inf. Process. Syst.*, 1990, pp. 396–404.

[12] D. Chen, X. Shao, B. Hu, and Q. Su, "Simultaneous wavelength selection and outlier detection in multivariate regression of near-infrared spectra," *Anal. Sci.*, vol. 21, no. 2, pp. 161–166, 2005.

[13] C. Spence, L. Parra, and P. Sajda, "Detection, synthesis and compression in mammographic image analysis with a hierarchical image probability model," in *Proc. IEEE Workshop Math. Methods Biomed. Image Anal. (MMBIA)*, Dec. 2001, pp. 3–10.

[14] D. Pokrajac, A. Lazarevic, and L. J. Latecki, "Incremental local outlier detection for data streams," in *Proc. CIDM*, Mar. 2007, pp. 504–515.

[15] J. Kittler *et al.*, "Domain anomaly detection in machine perception: A system architecture and taxonomy," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 36, no. 5, pp. 845–859, May 2014.

[16] F. Alegre, A. Amehraye, and N. Evans, "A one-class classification approach to generalised speaker verification spoofing countermeasures using local binary patterns," in *Proc. IEEE 6th Int. Conf. Biometrics, Theory, Appl., Syst. (BTAS)*, Sep. 2013, pp. 1–8.

[17] J. Komulainen, A. Hadid, and M. Pietikäinen, "Generalized textured contact lens detection by extracting bsif description from cartesian iris images," in *Proc. IEEE Int. Joint Conf. Biometrics*, Sep. 2014, pp. 1–7.

[18] Y. Ding and A. Ross, "An ensemble of one-class svms for fingerprint spoof detection across different fabrication materials," in *Proc. IEEE Int. Workshop Inf. Forensics Secur. (WIFS)*, Dec. 2016, pp. 1–6.

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

[20] G. Pan, L. Sun, Z. Wu, and S. Lao, "Eyeblink-based anti-spoofing in face recognition from a generic webcamera," in *Proc. ICCV*, 2007, pp. 1–8.

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

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

[23] M. De Marsico, M. Nappi, D. Riccio, and J.-L. Dugelay, "Moving face spoofing detection via 3D projective invariants," in *Proc. ICB*, Mar. 2012, pp. 73–78.

[24] X. Tan, Y. Li, J. Liu, and L. Jiang, "Face liveness detection from a single image with sparse low rank bilinear discriminative model," in *ECCV* (Lecture Notes in Computer Science), vol. 6316, K. Daniilidis, P. Maragos, and N. Paragios, Eds. Berlin, Germany: Springer, 2010, pp. 504–517.

[25] K.-C. Lee, J. Ho, M.-H. Yang, and D. Kriegman, "Visual tracking and recognition using probabilistic appearance manifolds," *Comput. Vis. Image Understand.*, vol. 99, no. 3, pp. 303–331, Sep. 2005.

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

[27] A. Anjos, M. M. Chakka, and S. Marcel, "Motion-based counter-measures to photo attacks in face recognition," *IET J. Biometrics*, vol. 3, no. 3, pp. 147–158, Jul. 2013.

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

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

[30] G. Kim, S. Eum, J. K. Suhr, I.-D. Kim, K. R. Park, and J. Kim, "Face liveness detection based on texture and frequency analyses," in *Proc. ICB*, Mar. 2012, pp. 67–72.

[31] W. R. Schwartz, A. Rocha, and H. Pedrini, "Face spoofing detection through partial least squares and low-level descriptors," in *Proc. IJCB*, Oct. 2011, pp. 1–8.

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

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

[34] D. Menotti *et al.*, "Deep representations for iris, face, and fingerprint spoofing detection," *IEEE Trans. Inf. Forensics Security*, vol. 10, no. 4, pp. 864–879, Apr. 2015.

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

[36] Q. Zhen, D. Huang, Y. Wang, and L. Chen, "LPQ based static and dynamic modeling of facial expressions in 3D videos," in *Biometric Recognition* (Lecture Notes in Computer Science), vol. 8232, Z. Sun, S. Shan, G. Yang, J. Zhou, Y. Wang, and Y. Yin, Eds. Cham, Switzerland: Springer, 2013, pp. 122–129.

[37] B. Jiang, M. F. Valstar, B. Martinez, and M. Pantic, "A dynamic appearance descriptor approach to facial actions temporal modeling," *IEEE Trans. Cybern.*, vol. 44, no. 2, pp. 161–174, Feb. 2014.

[38] S. R. Arashloo and J. Kittler, "Dynamic Texture recognition using multiscale binarized statistical image features," *IEEE Trans. Multimedia*, vol. 16, no. 8, pp. 2099–2109, Dec. 2014.

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

[40] G. Zhao and M. Pietikäinen, "Dynamic texture recognition using local binary patterns with an application to facial expressions," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 29, no. 6, pp. 915–928, Jun. 2007.

[41] J. Galbally, S. Marcel, and J. Fierrez, "Image quality assessment for fake biometric detection: Application to iris, fingerprint, and face recognition," *IEEE Trans. Image Process.*, vol. 23, no. 2, pp. 710–724, Feb. 2014.

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

[43] V. Ojansivu and J. Heikkilä, "Blur insensitive texture classification using local phase quantization," in *Image and Signal Processing* (Lecture Notes in Computer Science), vol. 5099, A. Elmoataz, O. Lezoray, F. Nouboud, and D. Mammass, Eds. Berlin, Germany: Springer, 2008, pp. 236–243.
[44] J. Kannala and E. Rahtu, "BSIF: Binarized statistical image features," in *Proc. ICPR*, Nov. 2012, pp. 1363–1366.
[45] S. Bayram, I. Avcıbaş, B. Sankur, and N. Memon, "Image manipulation detection," *J. Electron. Imag.*, vol. 15, no. 4, p. 041102, Dec. 2006.
[46] I. Avcibas, N. Memon, and B. Sankur, "Steganalysis using image quality metrics," *IEEE Trans. Image Process.*, vol. 12, no. 2, pp. 221–229, Feb. 2003.
[47] D. Tax, "One-class classification," Ph.D. dissertation, Delft Univ., Delft, The Netherlands, 2001. [Online]. Available: http://homepage.tudelft.nl/n9d04/thesis.pdf
[48] V. Vapnik, *Statistical Learning Theory*. Hoboken, NJ, USA: Wiley, 1998.
[49] B. Schölkopf, J. C. Platt, J. C. Shawe-Taylor, A. J. Smola, and R. C. Williamson, "Estimating the support of a high-dimensional distribution," *Neural Comput.*, vol. 13, no. 7, pp. 1443–1471, 2001.
[50] Z. Zhang, Y. Xu, J. Yang, X. Li, and D. Zhang, "A survey of sparse representation: Algorithms and applications," *IEEE Access*, vol. 3, no. 1, pp. 490–530, May 2015.
[51] J. Wright, A. Y. Yang, A. Ganesh, S. S. Sastry, and Y. Ma, "Robust face recognition via sparse representation," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 31, no. 2, pp. 210–227, Feb. 2009.
[52] D. L. Donoho and Y. Tsaig, "Fast solution of ℓ_1-norm minimization problems when the solution may be sparse," *IEEE Trans. Inf. Theory*, vol. 54, no. 11, pp. 4789–4812, Nov. 2008.
[53] M. Osborne, B. Presnell, and B. Turlach, "A new approach to variable selection in least squares problems," *IMA J. Numer. Anal.*, vol. 20, no. 3, pp. 389–403, 2000.
[54] C.-C. Chang and C.-J. Lin, "LIBSVM: A library for support vector machines," *ACM Trans. Intell. Syst. Technol.*, vol. 2, no. 3, pp. 27:1–27:27, 2011. [Online]. Available: http://www.csie.ntu.edu.tw/~cjlin/libsvm
[55] D. M. J. Tax and R. P. W. Duin, "Support vector data description," *Mach. Learn.*, vol. 54, no. 1, pp. 45–66, Jan. 2004.
[56] K. Patel, H. Han, and A. K. Jain, *Cross-Database Face Antispoofing with Robust Feature Represent.*. Cham, Switzerland: Springer, 2016, pp. 611–619. [Online]. Available: http://dx.doi.org/10.1007/978-3-319-46654-5_67
[57] K. Patel, H. Han, and A. K. Jain, "Secure face unlock: Spoof detection on smartphones," *IEEE Trans. Inf. Forensics Security*, vol. 11, no. 10, pp. 2268–2283, Oct. 2016.
[58] Z. Boulkenafet, J. Komulainen, and A. Hadid, "Face spoofing detection using colour texture analysis," *IEEE Trans. Inf. Forensics Security*, vol. 11, no. 8, pp. 1818–1830, Aug. 2016.
[59] J. Liu and A. Kumar, "Detecting sensor level spoof attacks using joint encoding of temporal and spatial features," in *Proc. 7th Int. Conf. Imag. Crime Detection (ICDP)*, Madrid, Spain, Nov. 2016, pp. 1–6.

SHERVIN RAHIMZADEH ARASHLOO received the Ph.D. degree in computer vision from the Center for Vision, Speech and Signal Processing, University of Surrey, U.K. He is currently an Assistant Professor with the Department of Medical Informatics, Faculty of Medical Sciences, Tarbiat Modares University, Tehran, Iran. His research interests include biometrics, graphical models, and cognitive vision.

JOSEF KITTLER is currently a Professor of machine intelligence with the Centre for Vision, Speech and Signal Processing, University of Surrey. He conducts research in biometrics, video and image database retrieval, medical image analysis, and cognitive vision. He published the book *Prentice Hall textbook on Pattern Recognition: A Statistical Approach* and over 170 journal papers. He serves on the Editorial Board of several scientific journals in pattern recognition and computer vision.

WILLIAM CHRISTMAS submitted his Ph.D. dissertation titled "Structural matching in computer vision using probabilistic reasoning." He was a Research Engineer with the BP International Research Center, Sunbury, U.K., and the BBC Engineering Research Department. He was also a University Research Fellow in technology transfer with the Centre for Vision, Speech and Signal Processing, University of Surrey, U.K. His interests included automated face analysis, with a particular interest in the use of 3-D morphable models, and the various aspects of sports video analysis and annotation.