Face Spoofing Detection Based on Multiple Descriptor Fusion Using Multiscale Dynamic Binarized Statistical Image Features

Shervin Rahimzadeh Arashloo and Josef Kittler, *Life Member, IEEE*

***Abstract*—Face recognition has been in the focus of attention for the past couple of decades and, as a result, a significant progress has been made in this area. However, the problem of spoofing attacks can challenge face biometric systems in practical applications. In this work, an effective countermeasure against face spoofing attacks based on a kernel discriminant analysis approach is presented. Its success derives from three innovations. First it is shown that the recently proposed multiscale dynamic texture descriptor based on binarized statistical image features on three orthogonal planes (MBSIF-TOP) is effective in detecting spoofing attacks, achieving promising performance compared to existing alternatives.**

Next, by combining MBSIF-TOP with a blur tolerant descriptor, namely the dynamic multiscale local phase quantization representation (MLPQ-TOP), the robustness of the spoofing attack detector can further be improved. The fusion of the information provided by MSIF-TOP and MLPQ-TOP is realized via a kernel fusion approach based on a fast kernel discriminant analysis (KDA) technique. It avoids the costly eigen-analysis computations by solving the KDA problem via spectral regression. The experimental evaluation of the proposed system on different databases demonstrates its advantages in detecting spoofing attacks in various imaging conditions, as compared to existing methods.

***Index Terms*—Face Spoofing, Multiscale Binarized Statistical Image Features on Three Orthogonal Planes (MBSIF-TOP), Multiscale Local Phase Quantization on Three Orthogonal Planes, Kernel Discriminant Analysis, Kernel Fusion.**

I. Introduction

ALTHOUGH face recognition technology has witnessed a significant progress, in the past couple of decades, from systems operating in well controlled laboratory settings to real world solutions for unconstrained scenarios, the operational utility of these systems can be challenged by artificial biometric traits, *i.e.* spoofing attacks. In a spoofing attack, an imposter tries to gain illegitimate access to some service by presenting artificial biometric data of another subject to the authentication system. In a recent study, it has been observed that face recognition systems are quite vulnerable to such attacks, as nearly 80% of the spoofing attempts successfully passed the authentication stage [1]. This vulnarability emphasizes the need for checking the authenticity of the biometric data before proceeding to verification or recognition.

Spoofing is not specific to face recognition systems. Other biometrics modalities suffer from similar drawbacks [2], [3], [4], [5]. However, the abundance of still face images or video sequences on the internet has made it particularly easy to access a person's facial data compared to other modalities. Moreover, the relatively low cost of launching a face spoof attack has made the face spoofing problem even more common. The media used for spoofing a face recognition system vary from low quality paper prints to high quality photographs, as well as video streams played in front of the biometric authentication system sensor. Other media such as 3D masks are less common [6].

Our focus in this paper is on attacks performed using a printed photograph or a replayed video in front of the system sensor. As spoofing attacks are realized using a variety of different media in a wide range of different imaging conditions, the problem poses serious challenges in practical applications. The research into face spoofing has become popular in the past few years, partly motivated by the recently organized competitions on countermeasures to 2D facial spoofing attacks [7], [8].

In the spectrum of spoofing countermeasure approaches in the literature, an important group of methods focuses on modeling the dynamic textural content of image sequences using spatio-temporal descriptors. An example of such methods is the work in [9] which employs local binary pattern histograms on three orthogonal planes (LBP-TOP) to detect spoofing attacks. The current work follows the same avenue and presents an effective method for the detection of spoofing attacks using a dynamic texture descriptor that is novel to this application domain. More specifically, motivated by the success of the BSIF descriptor [10] and its multiscale and dynamic extensions in a variety of static and dynamic texture representation and recognition problems such as face image modeling and recognition [11], dynamic texture recognition [12], finger print spoofing detection [13], *etc.*, the current work employs the dynamic multiscale BSIF descriptor [12] for face spoofing detection.

The BSIF descriptor operates in a similar fashion to the well known local binary pattern (LBP) operator in that it produces a binary coded image. However, as the BSIF descriptor employs filters based on statistical learning, it provides a better representation of image/image-sequence content. Most importantly, BSIF filters are designed to promote statistical independence

S.R. Arashloo is with the Faculty of Engineering, Urmia university, Urmia, Iran and also with the the Centre for Vision, Speech and Signal Processing (CVSSP), University of Surrey, Guildford, Surrey, UK, e-mail: Sh.Rahimzadeh@hotmail.co.uk, S.Rahimzadeh@surrey.ac.uk.

J. Kittler is with the Centre for Vision, Speech and Signal Processing (CVSSP), University of Surrey, Guildford, Surrey, UK.

Manuscript received ?; revised ?.

of their outputs. This enhances their representational capacity, in comparison with operators producing dependent outputs. Arguably this also improves their sensitivity to differences in the visual content of genuine and spoofing access attacks. A contribution of the current work is to employ the dynamic multiscale BSIF descriptor (MBSIF-TOP) [12] in conjunction with kernel discriminant analysis (KDA) for face aliveness assurance. The specific KDA method employed (SR-KDA) avoids costly eigen-analysis computations via spectral regression. It has been found to be orders of magnitude faster than the ordinary KDA [14].

A different but closely related dynamic texture descriptor is the local phase quantization (LPQ) representation [15]. The local phase quantization and its extension to time-varying textures (LPQ-TOP) [16], [17], which exploit the blur-insensitive property of the Fourier phase spectrum, have shown promise in texture/dynamic-texture modeling, especially when the acquired images suffer from blurring effects. In our study we include the LPQ descriptor because of its invariance to blur that may be particularly relevant to spoofing attack detection. We show that the LPQ-TOP representation is very effective in discriminating real access attempts from a certain type of spoofing attack in a subspace determined using the SR-KDA method.

In order to benefit from the complementary properties of both, MBSIF-TOP and MLPQ-TOP, we combine these two descriptors using the SR-KDA approach. The kernel fusion is found to enhance the system performance as measured on the CASIA face anti-spoofing database [18], as well as the Replay-Attack [1] and NUAA databases [19].

The main contributions of the current work can be summarised as follows:

- We propose the use of a discriminative representation based on the dynamic multiscale binarized statistical image features [12] and kernel discriminant analysis for face anti-spoofing. The representation is shown to perform better than similar dynamic texture descriptors such as LBP-TOP [20] and LPQ-TOP [16], [17]. An in-depth analysis of the properties of the BSIF descriptor that render it well suited for spoofing attack detection is beyond the scope of this paper. The key practical motivation is the promising performance that BSIF descriptors are shown to deliver in the context of the spoofing detection application.
- We advocate the spectral regression kernel discriminant analysis (SR-KDA) as a computationally efficient approach to implementing KDA.
- We demonstrate the beneficial impact of combining the novel descriptor (MSIF-TOP) with MLPQ-TOP on the robustness of the spoofing detection system. The kernel fusion is also accomplished by SR-KDA. On the CASIA dataset, the fusion strategy adopted is shown to be particularly effective in real-world scenarios such as low resolution biometric data or short image sequences. On the Replay-Attack and NUAA databases, the fusion consistently improves the performance.
- The proposed system has been evaluated in challenging conditions using standard benchmarking datasets and associated protocols. In addition, the dependence of its performance on data quality, on the type of spoofing media, and the number of video frames has been investigated. We also report the results of its comparison to similar dynamic texture descriptors and other face anti-spoofing approaches.

The remainder of the paper is organized as follows. Section II reviews the recent advances in face spoofing detection. Section III presents the proposed countermeasure along with details of the dynamic texture descriptors employed. The kernel discriminant analysis technique based on spectral regression and the kernel fusion method are also described. An experimental evaluation of the proposed method on different databases along with its comparison to other approaches is presented in Section IV. The paper is drawn to conclusions in Section V.

II. Related Work

A variety of different methods have been proposed for discriminating a real access from a fake one in a face authentication system. The research in this field has been reviewed in a recent study [9] where the various approaches to face liveness detection have been categorized according to the cues employed. These categories include the methods detecting signs of vitality (liveness), the methods gauging differences in motion patterns and those based on differences in image quality.

Examples of the methods of the first category invariably utilize characteristics which are exhibited only by live faces. For example, the method presented in [21] uses eye-blink as a countermeasure to spoofing, formulated as inference in an undirected conditional graphical framework. The method has been evaluated on a publicly available blinking video database [22]. However, eye-blink may not be considered as a reliable countermeasure as spoofing attacks using printed masks with the eye positions cut out can potentially pass such tests. Eye-blink in conjunction with other cues are employed in other works. For instance, in [23], a hybrid face liveness detection system against spoofing with photographs, videos, and 3D models is proposed. In this method, both eye-blink and scene context clues are utilized. Eye-blink detection is formulated as in [21], while the scene context clue is extracted by comparing the differences of regions of interest between the reference scene image and the input.

More recently, the authors in [24] proposed to use the dynamic mode decomposition (DMD) method for modeling dynamic information of the video content, such as blinking eyes, moving lips, and facial dynamics. The authors propose a classification pipeline consisting of DMD, local binary patterns and SVMs. The DMD approach is used to represent the temporal information of an image sequence as a single image. LBP operators are then applied on a single dynamic mode followed by an SVM for final decision making. As the method uses motion cues, it can also be considered as belonging to the second class in our categorization of anti-spoofing methods, discussed next.

The second group is comprised of the methods which assume that motion patterns between different components of

a real face differ from those of a fake face. This assumption is based on the fact that the spoofing media are usually flat 2D planes compared to the 3D structure of a real face. Moreover, motions in spoofing attacks are rigid whereas a combination of both rigid and non-rigid motions exists in real access attempts which may serve as a complementary cue for spoofing detection. As an example of the methods in this group, one may consider the work in [25] which presents a spoofing detection method for video sequences using motion magnification. For this purpose, Eulerian motion magnification is used to enhance the facial expressions. Two kinds of features are deployed: the first one is based on a configuration of LBP and the second is built upon the motion estimation approach using the HOOF descriptor. The method has been reported to achieve good performance on the Print Attack and Replay Attack data sets. Other work in [26], presented a method for evaluating liveness in face image sequences using a set of automatically located facial points. Geometric invariants are then used for detecting replay attacks. The proposed system was evaluated on two publicly available databases of NUAA [19] and HONDA [27].

In [28], based on the differences in optical flow fields generated by the movements of 2D planes and 3D objects, a new liveness detection method for face spoofing detection is proposed. Using the assumption that the test region is a 2D plane, a reference field from the actual optical flow field data is obtained. Next, a measure of difference between the two fields is used to distinguish a 3D face from a 2D photograph. Good performance has been obtained on a local data set. Other work in [29] proposed a countermeasure based on foreground/background motion correlation using optical flow. The method was shown to exhibit promising results on the publicly available Photo-Attack database.

The last category of face liveness detection methods relies on image quality measures to discriminate a live face from a fake one. These methods mainly rely on the reflectance properties of attack media which differ from those of a real face. As an example, the method in [30] proposes a multi-spectral face liveness detection method adaptive to various user-system distances. Using the Lambertian model, the multi-spectral properties of human skin versus non-skin are analyzed and the discriminative wavelengths are then chosen. An SVM classifier is then trained to learn the multi-spectral distribution for a final genuine/fake decision. Good performance has been reported on a private data set.

The work in [31] detects print-attacks by exploiting differences in the 2D Fourier spectra of spoofing attempts based on hard-copies of faces and real access attempts. The method is reported to work well for down-sampled photo attacks, however it would probably fail for higher-quality data. Based on the assumption that most of the information content of real images concentrates in specific frequency bands, the work in [18] used a set of difference of Gaussian filters to choose a specific frequency band, to be used as the features for discriminating real accesses from spoofing attempts. The method has been evaluated on the CASIA face anti-spoofing data set. Other work in [19] used the Lambertian model and proposed two methods to extract the essential information of different surface properties of a live human face or a photograph. The first one is a variational Retinex-based method while the second uses differences of Gaussian filters. Based on these, two extensions to the sparse logistic regression model were developed and evaluated on the NUAA data set.

Inspired by image quality assessment, the properties of printing artifacts and differences in light reflection that can be detected using texture features, in [32], it is proposed to detect spoofing from a texture analysis point of view. The approach analyses the texture of facial images using multiscale local binary patterns operators and is evaluated on the NUAA publicly available data set. The work in [33] proposes a single image-based face liveness detection method for discriminating 2D paper masks from the live faces. The method exploits frequency and texture information using power spectrum and LBPs. Three liveness detectors utilizing LBP, power spectrum and the fusion of the two were tested on two databases. In [34], an anti-spoofing method based on a set of low-level feature descriptors is proposed. The approach exploits both spatial and temporal information to learn distinctive characteristics of the class of real and spoof access attempts respectively. The descriptors employed encode information about shape, color and texture. The work in [35] studies the problem of fusion of motion and texture based methods and proposes a score level fusion to combine different visual cues to improve performance. The method has achieved good performance on the Replay-Attack database.

The authors of [36] propose to illuminate the face during image capture. The reflected color from the face served as a means to watermark the image. The method is based on the assumption that a pre-recorded video is highly unlikely to contain the correctly reflected color sequence. In [37], a face spoof detection algorithm based on image distortion analysis is proposed. Four different features of specular reflection, blurriness, chromatic moment, and color diversity are utilized to form an image distortion feature vector. An ensemble classifier, consisting of multiple SVMs trained for different face spoof attacks is then employed to discriminate genuine faces from spoofing attacks. Motivated by the success of deep networks, the work in [38] considers two approaches for face spoofing detection. The first approach investigates suitable convolutional network architectures while the second tries to learn the weights of the network via back-propagation. The method then combines and contrasts the two different approaches.

While most of the existing approaches are subject-independent, recently there have been some attempts to deploy subject-specific information in a face anti-spoofing system. For instance, the work in [39] proposes a subject-specific face anti-spoofing approach which tries to dismiss the interference among subjects using a classifier specifically trained for each subject. The work presents a subject domain adaptation method to handle the lack of fake samples during training and synthesizes virtual features to train individual face anti-spoofing classifiers. The idea is validated on the CASIA FASD and Replay-Attack data sets using MsLBP and HOG features.

Other work in [40], studied the client-specific information embedded within the features used and analyzed how it affects

the performance of a face anti-spoofing system. The authors built two anti-spoofing solutions, one relying on a generative and another one on a discriminative paradigm. Using both texture and motion cues, the proposed approach has been reported to be superior to the client-independent methods on the Replay-Attack data set.

III. Methodology

The proposed approach for face liveness detection utilizes histograms of dynamic texture descriptors on three orthogonal planes to encode texture micro-structure of an image sequence. For this purpose, the use of two effective spatio-temporal texture descriptors, namely histograms of multiscale dynamic binarized statistical image features (MBSIF-TOP) [12] and multiscale dynamic local phase quantization (MLPQ-TOP) [17], [16], [15] is advocated. Once the histograms of dynamic textures are obtained, they are projected onto a discriminative subspace constructed by KDA to separate real faces from spoofing attempts. The discriminative subspace for face liveness detection is constructed using an efficient kernel discriminant analysis approach based on spectral regression (SR-KDA) which has been found to be faster than the ordinary KDA by avoiding costly eigen-analysis computations [14]. The two representations are then combined to further enhance system performance. Their fusion is again accomplished using the SR-KDA method. The multiple kernels corresponding to different representations are combined via a sum rule over the kernel matrices.

The proposed kernel fusion technique for face spoofing detection improves the performance, compared to either one of the descriptors used individually, in most cases. It is computationally more efficient compared to conventional KDA methods. This makes the proposal a viable solution as a preliminary step in checking the authenticity of the biometric data captured by a face biometric system. A description of the two descriptors followed by an outline of the SRKDA method is provided next.

A. Descriptors

1) Binarized Statistical Image Features(BSIF): The binarized statistical image features are the key constituent of a generative texture descriptor based on independent component analysis [10]. The BISF descriptor uses pre-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

$$s_i = \sum_y W_i(y)X(y) = w_i^\top x \tag{1}$$

where vectors x and w_i contain the pixels of X and elements of W_i, respectively while $.^\top$ denotes transpose. The binarized feature b_i is then obtained by thresholding the filter response s_i at zero, *i.e.*

$$b_i = \begin{cases} 1 & s_i > 0, \\ 0 & \text{otherwise.} \end{cases} \tag{2}$$

In summary, given an image patch X of size $l \times l$ pixels, one applies N filters to X using the filter matrix $W^{N \times l^2}$ and obtains N responses which are stacked into vector $\mathbf{s}$. The filter responses are independently binarized next to form an N-bit binary codeword for each pixel which is then converted into an integer number.

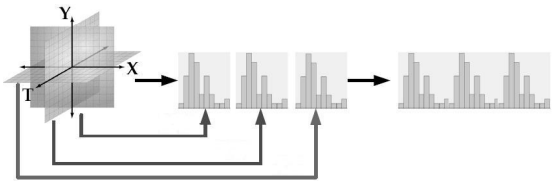

Fig. 1. Construction of a histogram on three orthogonal planes in the MBSIF-TOP, MLPQ-TOP and MLBP-TOP representations.

Learning BSIF Filters:

In order to obtain a useful set of filters, the statistical independence of s_i's is maximized using independent component analysis (ICA). For this purpose, the filter matrix W is decomposed into two parts as

$$s = Wx = UVx = Uz \tag{3}$$

where $z = Vx$, and U is an $N \times N$ square matrix estimated via ICA. Matrix V performs simultaneous whitening and dimensionality reduction using principal component analysis. Given whitened data samples z, the independent component analysis is then used to estimate an orthogonal matrix U. Finally, given U and V, one obtains the filter matrix $W = UV$. In this work, 8 filters are used ($N = 8$). Once the 8 filter responses are obtained at all pixel locations and the corresponding 8-bit codewords are generated and converted to integer numbers, a histogram is used to summarize the filter outputs.

In [12], the BSIF approach is extended to the spatio-temporal domain by considering image sequence on three orthogonal planes, similar in spirit to the LBP-TOP approach [20]. In the LBP-TOP approach [20], an image sequence is considered as a stack of XT planes in axis Y, YT planes in axis X and XY planes in axis T, where the YT and XT planes convey information about the space-time transitions of a video sequence while the XY plane represents spatial information. In order to extend the BSIF descriptor to the spatio-temporal domain the same procedure as in the static case is followed to learn three different sets of BSIF filters on different planes. Once the filters are applied to each individual plane, the filter responses are summarized using histograms. Finally, the three histograms thus obtained are normalized to yield a probability distribution and concatenated to form the final BSIF-TOP descriptor, Fig. 1. In [12], the BSIF-TOP approach is also extended to a multiscale framework to capture dynamic content at multiple resolutions. The extension of the BSIF-TOP to a multiscale framework entails applying filters of various sizes to an image. In this work, the filters are learned in six scales using an external set of image sequences of random scenes using the fast ICA package [41].

2) Local Phase Quantization(LPQ): The local phase quantization (LPQ) is a texture representation approach based on the blur invariance property of the Fourier phase spectrum [15]. The LPQ descriptor uses local phase information extracted using a short-term Fourier transform computed over a

rectangular region. The short-term Fourier transform over a region of size $l \times l$ centered at pixel position m of the image $X(y)$ is defined as

$$F(f,m) = \sum_{y} X(m-y)e^{-\mathrm{j}2\pi f^{\top} y} = w_f^{\top} x \tag{4}$$

where w_f denotes the basis vector of the 2D discrete Fourier transform at frequency f while x stands for the vector containing all l^2 pixels. In the LPQ representation, the local Fourier coefficients are computed at four frequencies $f_1 = [a,0]^{\top}$, $f_2 = [0,a]^{\top}$, $f_3 = [a,a]^{\top}$ and $f_4 = [a,-a]^{\top}$, where a is a sufficiently small scalar. The result for each pixel location is a vector $F_m^C = [F(f_1,m), F(f_2,m), F(f_3,m), F(f_4,m)]$. Assume

$$F_m = [Re\{F_m^C\}, Im\{F_m^C\}] \tag{5}$$

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_m as $F_m = Tx$ would then be

$$T = [Re\{w_{f_1}, w_{f_2}, w_{f_3}, w_{f_4}\}, Im\{w_{f_1}, w_{f_2}, w_{f_3}, w_{f_4}\}]^{\top} \tag{6}$$

Let the covariance matrix of the transform coefficient vector F_m be D. A singular value decomposition of matrix D is obtained as

$$D = A\Sigma B^{\top} \tag{7}$$

The matrix B is then used to perform a whitening transform on F_m as

$$G_m = B^{\top} F_m \tag{8}$$

Once G_m is computed for all pixel positions, the information in the Fourier coefficients is recorded by binarizing the elements of G_m as

$$q_j = \begin{cases} 1 & \text{if } g_j \geq 0, \\ 0 & \text{otherwise.} \end{cases} \tag{9}$$

where g_j is the j^{th} component of the vector G_m. The resultant 8-bit binary coefficients q_j are then represented as integers. Once the binary codes are obtained, a histogram of the obtained values from all positions is constructed.

The same approach as in the BSIF-TOP [12] and LBP-TOP [20] methods is employed to extend the LPQ descriptor to the temporal domain [16], [17], *i.e.* the basic LPQ features are extracted independently from the sets of three orthogonal planes: XY, XT and YT, considering only the co-occurrence statistics in these three directions, and stacking them into a single histogram. In order to capture texture details at different scales, the window sizes are varied and in total six different window sizes of $\{3 \times 3, 5 \times 5, \ldots, 13 \times 13\}$ are employed in this work to obtain a multiscale representation of dynamic textural content of an image sequence. The histograms of each plane are normalized independently so that each sums to 1 to yield a coherent representation. They are then concatenated to form the final MLPQ-TOP descriptor.

In the current work, all the frames of given a video sequence are used for feature extraction from the XY, XT and YT planes. The discrepancies between different video sequences in terms of number of frames are compensated for by normalizing the histograms obtained. The effect of using video of different lengths on system performance is investigated in Section IV.

3) Motivation for the advocated descriptors: Foremost, the list of motivations in support of the proposed descriptors for the design of a face biometric spoofing detector includes the desirable properties of BSIFs which are inherited from the BSIF filters being designed so as to promote statistical independence of their outputs. This enhances their representational capacity, as compared with operators producing dependent outputs. Arguably this also improves their sensitivity to subtle differences in the visual content of genuine and spoofing access attacks. Moreover, as BSIF filters can be learned using an independent set of random image sequences distinct from the databases used for evaluation, the BSIF representation circumvents the need for application specific filter design and parameter tuning. As such, the BSIF representation may act as a generic descriptor to handle different kinds of face artifacts in real-life situations. On the other hand, the LPQ descriptor is known to possess invariance to blurring effects which may be relevant to spoofing attack detection. In addition, employing the aforementioned descriptors in a multiscale framework helps to capture highly varying face information more effectively, and this appears to enhance robustness of the system.

B. Kernel Discriminant Analysis Using Spectral Regression

Assume that there exist m samples $x_1, x_2, \ldots x_m \in \mathbb{R}^n$, which belong to C classes and $\mathcal{F}$ is a feature space induced by a non-linear mapping $\phi : \mathbb{R}^n \rightarrow \mathcal{F}$. For a suitably chosen mapping, an inner product $\langle ., . \rangle$ on $\mathcal{F}$ may be represented as $\langle \phi(x_i), \phi(x_j) \rangle = \kappa(x_i, x_j)$, where $\kappa(.,.)$ is a positive semi-definite kernel function. Let S_b^{ϕ} and S_t^{ϕ} denote respectively the between-class and total scatter matrices in the feature space $\mathcal{F}$. KDA seeks to find an optimal projection function V_{opt} in the feature space by solving the following optimization problem

$$V_{opt} = \arg\max_{V} \frac{V^{\top} S_b^{\phi} V}{V^{\top} S_t^{\phi} V} \tag{10}$$

The columns of V_{opt} (ν's) are the generalized eigenvectors satisfying

$$S_b^{\phi} \nu = \lambda S_t^{\phi} \nu \tag{11}$$

It is known that ν's satisfying the preceding problem can be expressed as linear combinations of all samples [42], [43]. Thus, there exists coefficients α_i such that each eigenvector ν can be represented as $\nu = \sum_{i=1}^{m} \alpha_i \phi(x_i)$.

In [42], it is shown that Eq. 10 is equivalent to

$$U_{opt} = \arg\max_{U} \frac{U^{\top} KWKU}{U^{\top} KKU} \tag{12}$$

where K is the kernel matrix ($K_{ij} = \kappa(x_i, x_j)$) and W is a matrix reflecting the number of samples in each class, defined as

$$W_{ij} = \begin{cases} 1/m_k & \text{if } x_k,\ x_j \in k^{th} \text{ class;} \\ 0 & \text{otherwise.} \end{cases} \tag{13}$$

In this case, the columns of U_{opt} (α's) are given by the eigenvectors corresponding to the non-zero eigenvalues satisfying

$$KWK\alpha = \lambda KK\alpha \tag{14}$$

The number of α's satisfying Eq. 14 is bounded by $C-1$ as the rank of S_b^ϕ is at most $C-1$. Once α's are found, the projection of a new sample (x) onto the feature space using each eigenvector ν is then given by

$$\begin{aligned}\langle \nu, \phi(x) \rangle &= \sum_{i=1}^{m} \alpha_i \langle \phi(x_i), \phi(x) \rangle = \sum_{i=1}^{m} \alpha_i \kappa(x_i, x) \\ &= \alpha^\top K(:, x) \end{aligned} \tag{15}$$

where $K(:,x) = [\kappa(x_1, x), \ldots, \kappa(x_m, x)]^\top$.

It is shown in [14] that instead of solving Eq. 12 via the usual eigen-analysis computations, the KDA projections can be obtained using spectral regression by solving the following two linear equations:

1) Solve $Wy = \lambda y$ for y;
2) Find α satisfying $K\alpha = y$.

where y is an eigenvector of W. The Cholesky decomposition is used to solve the linear equation $K\alpha = y$. If K is positive-definite, then there exists a unique solution for α. If K is singular, it may be approximated by the positive definite matrix $K + \delta I$ where I is the identity matrix and $\delta > 0$ is a regularization parameter. In this paper, a Gaussian RBF is used for the kernel function, *i.e.* $K_{ij} = \kappa(x_i, x_j) = \exp(-\|\mathbf{x}_i - \mathbf{x}_j\|^2/M)$, resulting in a positive definite kernel matrix [14], [43]. Following [44], [45], M is set to the average squared Euclidean distance between all training samples. Solving for α may then be performed using the Cholesky factorization and forward-back substitution. Finding vector y is trivial and can be performed directly without eigen-analysis [14]. As a result, one only needs to solve a set of regularized regression problems and there is no eigenvector computation involved which allows handling large kernel matrices and results in a considerable reduction of computational cost [14].

1) Kernel Fusion: In our multiple descriptor-based approach, there are two sources of information, each provided by a different descriptor which can be combined to enhance system performance. Although other alternatives exist [46], linear kernel combinations [47] are amongst the widely employed kernel fusion techniques. In this work, we have opted for a linear combination approach with a uniform weighting assigned to the kernels associated with different representations. This choice is driven by the simplicity of the approach as well as the proved efficacy in other classification problems [11]. The motivation for fusion could be not only performance, but also robustness, *i.e.* even if in certain scenarios BSIF and LPQ can deliver excellent performance on their own, they are not the individually best descriptors in all scenarios. Fusion provides much more robust performance overall.

Once the kernels are combined, an optimal solution is found in the implicit feature space which merges the employed representations. This is accomplished by solving the problem in Eq. 12 with the kernel matrix replaced by K_c given as

$$K_c = K_B + K_L \tag{16}$$

where K_B and K_L correspond to the kernel matrices constructed using the MBSIF-TOP and MLPQ-TOP descriptors, respectively.

C. Testing A New Face

During testing, the combined kernel vector k_c is formed by

$$k_c = k_B + k_L \tag{17}$$

where the elements of the vectors k_B and k_L are the respective similarities (measured using the employed Gaussian RBF $\kappa(.,.)$) of a query face sequence and training samples using the MBSIF-TOP and MLPQ-TOP representations. As in the spoofing detection problem there are two classes, only a single transformation vector for the KDA projection would be obtained, *i.e.* only a single vector α will satisfy Eq. 14. In this case, after computing the inner product of the test vector and the unique α satisfying Eq. 14, the projection is compared against the projection of the mean of the positive training samples in the induced feature space (ω) and the distance thus obtained is used as a dissimilarity criterion. In this work, we have chosen the mean of the positive class (real accesses) as the reference point to measure the distance of a query pattern due to the fact that the mean of the imposter set (spoofing attacks) may not serve as a stable and suitable reference due to the diversity of the attack media. The architecture of the proposed system for face spoofing detection is given in Fig. 2.

IV. Experiments

In this section, an experimental evaluation of the proposed method on the CASIA Face Anti-Spoofing Database [18], the Replay-Attack database [1] and the NUAA database [19] is provided.

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

The performance of an algorithm may vary depending on the imaging quality. In this respect, one of the appealing properties of the CASIA FASD [18] is the inclusion of various imaging qualities. Three different cameras are used in this database to collect the data. The data set consists of video sequences of both real and fake access attempts. During recording, subjects were asked to blink and not to keep still. Three kinds of fake face attacks designed in this database are as follows.

Warped photo attack: A Sony NEX-5 camera was used to record a 1920×1080 image for every subject. This high resolution image was then used to print a photo on the copper paper, having a higher quality than normal A4 printing paper. In a warped photo attack, the attacker warps the photo, trying to simulate facial motion.

Cut photo attack: The photos mentioned above are then used for the cut photo attacks. In this scenario, the eye regions are cut off. An attacker then hides behinds the photo and blinks through the holes of the eye region. Another possibility is to place an intact photo tightly behind the cut photo, and simulate blinking by moving the photo behind. In this database, both implementations exist.

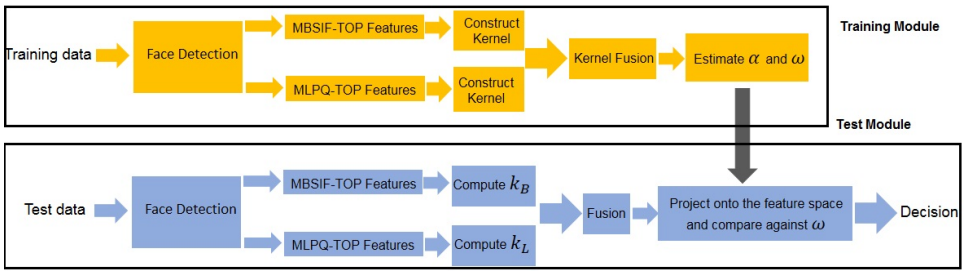

Fig. 2. Architecture of the proposed system.(α,ω, k_B and k_L are explained within the text.)

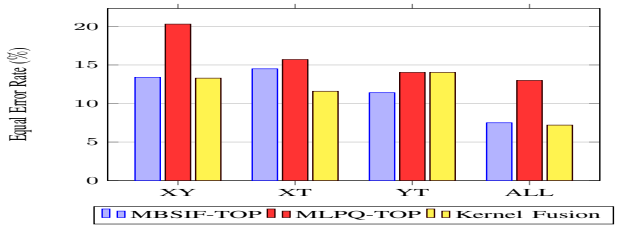

Fig. 3. Effectiveness of Different Planes of the employed dynamic representations on the CASIA FASD.

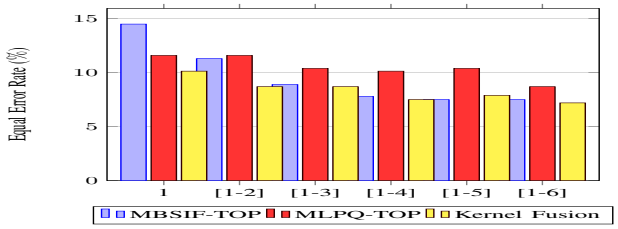

Fig. 4. Effect of Number of Scales Used in the Dynamic Descriptors on the CASIA FASD.

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

For CASIA FASD a test protocol is designed which consists of seven scenarios, including the warped photo test, cut photo test, video test and the quality and overall tests described next.

Quality Test: In this case, the three imaging qualities are considered separately. Specifically, the samples used for training and test are from low quality, normal quality and high quality video streams.

The Overall Test: In this scenario, all the data are combined together to be used for a general and overall evaluation.

The CASIA FASD is split into the training set comprised of 20 subjects and the test set containing 30 individuals. For each of the above seven test scenarios, the data should then be selected from the corresponding training and the test sets for model training and performance reporting. Detection-Error Trade-off (DET) curves [48] are utilized to evaluate the system accuracy. From DET curves, the point where the False Acceptance Rate (FAR) equals the False Rejection Rate (FRR) is located, and the corresponding value, *i.e.* the Equal Error Rate (EER) is reported.

In the following, we investigate the merits of the proposed approach in various conditions. In particular, the effectiveness of each of the three different planes in the the spatio-temporal multiscale representations, the effect of using a multiscale dynamic representation, the effect of quality and spoofing media and the effect of video sequence length on performance are investigated. Finally, a comparison to other existing methods is performed.

1) Effectiveness of Different Planes: First, the usefulness of spatial and temporal information is investigated. For this purpose, the MBSIF/MLPQ histograms on each of the three XY, XT and YT planes are extracted and used individually and also in combination for spoofing detection. The spectral regression kernel discriminant analysis (SRKDA) approach [14] is used to construct a discriminative representation for each of the four scenarios. The SRKDA approach is trained using all the real access image sequences and the videos of spoofing attacks of the training subset of the CASIA data set. For each video stream, the face is located using the Viola-Jones face detector. For this purpose, the detector is run on the first frame of each video sequence and the face location obtained is then generalized to all frames. In case the face detector fails to find a face in the first frame, the subsequent frame(s) is employed till a face is detected.

After resizing the detected face to 120×100 pixels, the multiscale BSIF and LPQ histograms are extracted on the XY, XT and YT planes. For performance evaluation, as indicated in the CASIA FASD protocol, the equal error rate (EER) on the test set is recorded. The results obtained are given in Fig. 3. As can be seen from the figure, the XT and YT planes indeed provide quite useful discriminative information. Interestingly, the spatio-temporal information on these planes proves to be as good or even better than the spatial information alone for classification. This can be observed for example by the lower EER obtained using the MBSIF-TOP representations on the YT plane compared to the XY plane. It can also be observed that using a histogram concatenation for fusing spatial and temporal information, the performance is improved for all three systems (BSIF, LPQ, and the fusion of the two descriptors). It can be concluded that the inclusion of dynamic information results in improved performance for face liveness detection.

2) Effect of a Multiscale Representation: Next, we investigate the merits of a multiscale dynamic representation for face spoofing detection. The effect of using a larger number of scales is investigated by gradually increasing the number of scales used. In this experiment, the $[x-y]$ notation denotes that all the scales between the x and y including the x and y scales are combined to form a multi-resolution representation. All the

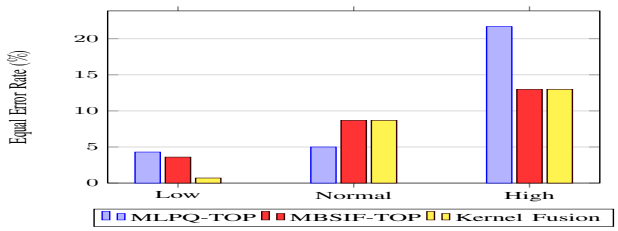

Fig. 5. Effect of Quality on the Performance on the CASIA FASD.

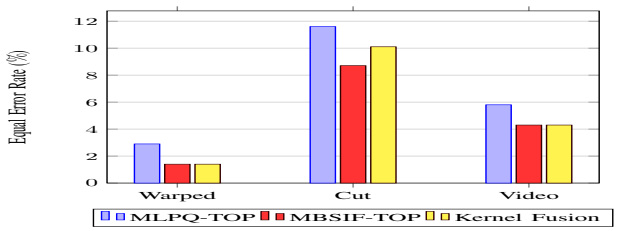

Fig. 6. Effect of Spoofing Media on the performance on the CASIA FASD.

positive and negative samples in the CASIA FASD are used to train the classifier and the performance is recorded as the EER on the test set. The results obtained on the CASIA FASD are illustrated in Fig. 4. As can be observed from the figure, apart from some fluctuations in EER's (especially for the LPQ representation), a multiscale representation generally performs better than a single scale approach. In the experiments to follow, the number of scales is fixed to six, as using six scales has been found almost to saturate the performance.

3) Effect of Quality and Spoofing Media: As noted earlier, in the CASIA FASD images corresponding to different qualities exist. These are low quality, normal quality and high quality image sequences. In addition, different media are used for spoofing attacks. The media used for spoofing attacks are warped paper, cut paper and the videos played on an iPad. In this section, we analyse the effects of quality of image sequences and spoofing media on the system performance. In this experiment, the training and test data are selected from the relevant data in the training and the test subsets. Three different qualities (low, normal and high) and three different media (warped photo, cut photo and video attacked) are examined separately.

The results of the analysis for the MBSIF-TOP, MLPQ-TOP and the kernel fusion methods are given in Fig. 5. In this experiment, six scales and a combination of three planes are used. As can be seen from the figure, with low quality image sequences, the proposed systems achieve relatively low equal error rates. With increasing image quality, the performance of all three systems degrades. The hypothesis is that one of the discriminating factors between the spoofing attacks and real accesses is the high frequency content of image sequences which are likely to be attenuated in spoofing attacks. However, as the high frequency content of spoofing attacks is strengthened by increasing the device quality, the ability to distinguish them from genuine accesses is diminished. Further investigation is needed to fully characterize the effect of quality on performance. It is pertinent also to examine the possibility of degrading the quality artificially, or deliberately using a poor quality device, to achieve better detection performance.

Regarding the effect of spoofing media, examination of Fig. 6 reveals that all three systems perform well for the warped attacks. However, as expected, the error rates increase when the spoofing media are face prints, in which eyes positions are cut out and blinking is performed. Fig. 7 depicts typical MBSIF-TOP histograms corresponding to cut-photo and warped-photo attacks for a subject from the CASIA FASD.

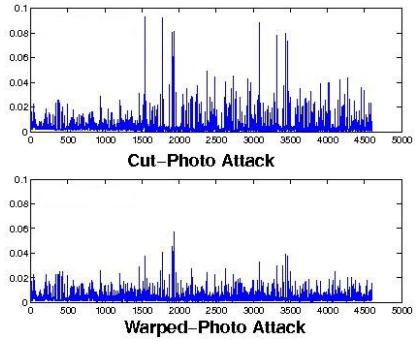

Fig. 7. MBSIF-TOP histograms corresponding to the cut-photo and warped-photo attack scenarios of a subject from the CASIA FASD.

Comparison of the two histograms reveals that the histogram corresponding to cut-photo attack has more dynamic content, partly caused by blinking. This is especially observable in the XT and YT components of the histograms, located on the right of the histograms which makes the features appear more similar to the features corresponding to real access attempts.

It can be concluded that blinking plays an important role for discriminating the spoofing attacks from real accesses. In the video attack scenario, the performance of the three systems based on BSIF, LPQ, and fusion deteriorates compared to the warped paper attack scenario. However, the error rates for the video attacks are better than those for the cut paper scenario which can be explained by the limited quality of the video displays in addition to reflections from an iPad screen making it easier to discriminate an attack from a real access request.

4) Effect of Time Window Length: In the previous experiments, all the frames of a relevant video sequence in the CASIA FASD [18] were used to construct the spatio-temporal representations. The video sequences in the CASIA FASD may be as long as 10 seconds. However, in practical applications such a long capture time should preferably be avoided to reduce the computational cost and speed up the

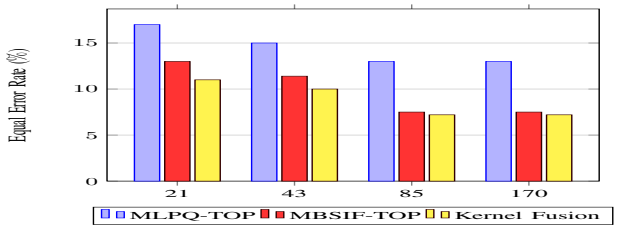

Fig. 8. Effect of Time Window Length on the performance on the CASIA FASD.

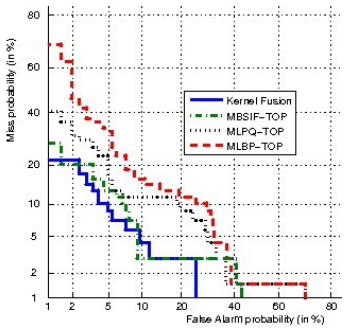

Fig. 9. DET curves of different dynamic representations in spoofing detection on the CASIA FASD.

whole process in the operation phase. In addition, long video sequences may not be available during training. In this section, the effects of video length (in terms of number of frames) on the performance is analyzed. For this purpose, the equal error rates corresponding to different video lengths are recorded. The results on the CASIA FASD are given in Fig. 8. From the figure, it can be observed that by increasing the number of frames, which results in a more stable representation, the error rates tend to decrease. However, the proposed systems' performance deteriorates slowly when using fewer frames. This is apparent from the equal error of $\approx 10\%$ obtained using less than two seconds of video (43 frames on average) for the kernel fusion approach. Moreover, reducing the length of the video sequences by half (from 170 to 85 frames on average) results in almost no loss of performance in neither one of the three systems.

5) Comparison of the MBSIF-TOP, MLPQ-TOP and MLBP-TOP: An experiment is carried out to compare the individual potential of MBSIF-TOP and MLPQ-TOP with that of the baseline descriptor, namely MLBP-TOP in detecting spoofing attacks. The three descriptors are trained using the training subset of the CASIA FASD and tested on the test subset. The results are reported as detection-error-trade off (DET) curves in Fig. 9. It can be observed from the figure that MBSIF-TOP and MLPQ-TOP descriptors perform better than the MLBP-TOP representation. Although MLPQ-TOP performs only slightly better than MLBP-TOP, MBSIF-TOP outperforms both descriptors by a larger margin. The performance of the system based on the fusion of MBSIF-TOP and MLPQ-TOP representations is also included for comparison.

6) Comparison of the Proposed Method to Other Approaches: Finally, we compare the three systems (MBSIF-TOP, MLPQ-TOP and the kernel fusion approach) in the seven scenarios of the CASIA FASD to other existing approaches reported in the literature. The results are reported in Table I. The following observations from the table are in order. First of all, in most cases, the MBSIF-TOP approach achieves better performance compared to the MLPQ-TOP. Most importantly, in the overall test scenario, the EER of the MBSIF-TOP is lower than that of the MLPQ-TOP which demonstrates the efficacy of the MBSIF-TOP representation for spoofing detection. Second, the kernel fusion of the two descriptors performs better or on par with either one of the single descriptors in five out of seven test scenarios. As already observed in the previous experiments, the fusion technique is especially advantageous when dealing with relatively short or low resolution video sequences. Third, the proposed approach based on a kernel fusion of MBSIF-TOP and MLPQ-TOP descriptors compares very favorably to the existing approaches. It can be observed that the kernel fusion technique achieves the lowest error rate in the overall test scenario among other competitors. In addition, in four out of the six remaining test scenarios the proposed fusion approach achieves the lowest error rates reported.

B. The Replay-Attack database

The Replay-Attack database [1] consists of 1200 short video recordings of both real-access and attack attempts of 50 different individuals. The data set includes 200 real access videos, 200 print attack videos, 400 phone attack videos, and 400 tablet attack videos. For performance evaluation, the data set is divided into three subsets of training (360 videos), development (360 videos), and testing (480 videos). The training and the development subgroups contain 60 real access videos and 300 attack videos each, whereas the testing subset contains 80 real access and 400 attack videos. The training set is used to train a classifier and the development set is typically used to adjust classifier parameters. The test set is intended to be used only for performance evaluation.

In the evaluation of the proposed method on this data set, we used the training set to learn the KDA projections while the development set is used to a set a threshold so that the false acceptance error rate is equal to the false rejection error rate. The threshold thus obtained is then used to classify a test pattern. In this experiment, we have used multiscale representations of the two descriptors on three orthogonal planes to form the features.

The results obtained on this data set along with some other methods are reported in Table II. A number of observations from the table are in order. First of all, it can be observed that the MBSIF-TOP descriptor performs better than the MLPQ-TOP representation. Second, the proposed anti-spoofing technique based on kernel fusion is instrumental in improving the classification accuracy. In this case, the half total error rate obtained on the test set of the replay-attack database using the fusion technique is 1.0% whereas the half total error rate obtained using the MBSIF-TOP representation is 1.38% and the corresponding error rate using the MLPQ-TOP histograms is 3.75%. The best reported result among others is 0.0%, obtained by both methods of CASIA and LNMIIT [8]. However, as also observed in [24] performance on the Replay-Attack database without properly normalizing face images may be artificially boosted. This is due to the fact that spoofing attacks in this data set have larger faces compared to the faces of the valid accesses. Moreover, in the proposed method only the face region, as detected by a generic face detection method, is utilized for spoofing detection whereas the CASIA and the LNMIIT methods [8] either rely on the background information or use different human body regions for classification.

TABLE I
COMPARISON OF EER'S (%) OF DIFFERENT METHODS ON THE CASIA FASD.

Scenario	Low	Normal	High	Warped	Cut	Video	Overall
MLPQ-TOP+KDA	4.3	**5**	21.7	2.9	11.6	5.8	13
MBSIF-TOP+KDA	3.6	8.7	13	1.4	8.7	4.3	7.5
Kernel Fusion	**0.7**	8.7	**13**	**1.4**	10.1	**4.3**	**7.2**
CASIA Baseline [18]	13	13	26	16	**6**	224	17
LBP-TOP [9]	10	12	13	6	12	10	10
IQA-based [49]	31.7	22.2	5.6	26.1	18.3	34.4	32.4

TABLE II
COMPARISON OF HALF TOTAL ERROR RATES (%) OF DIFFERENT METHODS ON THE EVALUATION AND TEST SETS OF THE REPLAY-ATTACK DATABASE.

Method	Development Set	Test Set
MLPQ-TOP+KDA	5.0	3.75
MBSIF-TOP+KDA	1.67	1.38
Kernel Fusion	**1.67**	**1.0**
$LBP^{u2}_{3\times3}$ [1]	14.84	15.16
$LBP + SVM$ [50]	13.9	13.87
$LBP - TOP + SVM$ [50]	7.88	7.60
IQA-based [49]	-	15.2
LLR [35]	4.57	5.11
Spoofnet [38]	-	3.5
IDA [37]	-	7.41
DMD+SVM [24]	8.5	7.5
DMD+LBP+SVM [24]	5.33	3.75
MsLBP [39]	0.87	1.45
HOG [39]	0.24	3.58
client-specific LBP-TOP [40]	3.71	3.95
HOOF+LDA (NN) [25]	0.0	1.25
CASIA [8]	0.0	0.0
LNMIIT [8]	0.0	0.0

TABLE III
COMPARISON OF EER'S (%) OF DIFFERENT METHODS ON THE NUAA DATA SET.

Method	**EER**
Gabor [32]	9.5
LPQ [32]	4.6
LBP[32]	2.9
The method of [51]	1.9
MLPQ-TOP	3.5
MBSIF-TOP	2.3
Kernel Fusion	**1.8**

C. The NUAA photograph imposter database

In the previous experiments, we have evaluated the system performance on video sequences. However, the spoofing/real access data may only be available in the form of still images. In this section, an analysis is performed to investigate the applicability of the system based on static MBSIF and MLPQ representations combined via the utilized fusion technique to detect spoofing attacks. The system architecture is exactly the same as in the previous experiments except that instead of dynamic MBSIF and MLPQ representation, their static versions operating on still images are used. A characteristic of the NUAA database [19] is that it consists of still images. The NUAA photograph imposter data set is collected using several generic cheap webcams [19]. The database is collected in three different sessions with about two weeks interval between two sessions, with the place and illumination conditions of each session being different.

In total, 15 subjects are included in the database. In each session, images of both live subjects and their photographs are captured. For each subject in each session, a series of images (with a frame rate of 20 fps and 500 images for each subject) are collected. During image capture time, each subject was asked to face the webcam frontally with neutral expression and with no apparent movements such as eye-blink or head movement. In other words, a live human tried to create a sequence of frames looking as if captured from a photo as much as possible. A high definition photo for each subject using a common Canon camera was captured in a way that the face area took at least 2/3 of the whole area of the photograph. The photos used for the attacks were then generated in two different ways. First, a photograph was printed on a photographic paper with the common size of $6.8cm \times 10.2cm$ (small) and $8.9cm \times 12.7cm$ (big). Next, each photo was printed on a 70g A4 paper using a usual color HP printer. Five categories of the photo-attacks are simulated in front of the webcam. Six scales of BSIF and LPQ features are used in this experiment. The performance of the proposed approach along with some other methods is reported in Table III. It can be seen that the kernel fusion technique compares very favorably to the other existing methods, ranking first in performance on this data set closely followed by the method of [51].

D. Cross-database Evaluation

A desirable characteristic of a classification system is its generalization capability. This aspect of the proposed method has been investigated on the Replay-Attack data set in an intra-database setting where the classifier was trained on a disjoint set and the performance was gauged on separate evaluation and test sets. The generalization ability of the system in this case is manifested by the relatively close error rates obtained on the evaluation and test sets (Table II). However, a more challenging scenario may be considered as using different databases for training and performance reporting. In order to investigate this aspect of the system we use the training subset of the Replay-Attack database to train the classifier. Next, the system is evaluated on the test set of the CASIA FASD. Given that the subjects and the imaging conditions and devices are different between the two databases, the test is quite challenging.

TABLE IV
COMPARISON OF EER'S (%) ON THE CASIA FASD USING CROSS-DATABASE AND ORIGINAL PROTOCOLS IN THE OVERALL SCENARIO.

Method	Original Protocol	Cross-database Evaluation
MLPQ-TOP	13	33.0
MBSIF-TOP	7.5	31.9
Kernel Fusion	7.2	30.2

The result of this evaluation in the overall scenario of the CASIA FASD is reported in Table IV. For comparison the results obtained on the CASIA FASD following an intra-database protocol (original protocol) are also reported. The results demonstrate a high degradation in performance indicating some degree of over-fitting of the classifier to data. One solution to the problem is to employ additional cues besides dynamic texture descriptors. Moreover, as proposed in [52] different strategies can be adopted to improve performance in such cross-database evaluations using the texture based representations. In a nearly similar cross-database setting in [52] the equal error rate obtained on the CASIA FASD was reported to be more than 60%, nearly twice the error rate obtained using the proposed method.

V. CONCLUSION

The paper addressed the problem of face biometrics spoofing detection based on dynamic texture analysis of biometric access video sequences. The use of the multiscale binarised statistical image features descriptor (MBSIF-TOP), a novel descriptor in the context of this application, was proposed as the representation to be adopted in a spoofing detection system designed using the computationally efficient spectral regression kernel discriminant analysis (SR-KDA). The proposed method was comprehensively evaluated on benchmarking datasets using standard protocols and shown to be superior to similar dynamic texture analysis schemes employing the MLBP-TOP and MLPQ-TOP descriptors. Its superior spoofing detection performance is attributed to the filters producing BSIF, which are designed to promote statistical independence of their outputs. The benefit of statistical independence is twofold: improved representational capacity of the texture descriptor, and its enhanced sensitivity to minor differences in the visual content of genuine and spoofing attack accesses.

A further performance improvement has been achieved by combining the MBSIF-TOP and MLPQ-TOP descriptors using kernel fusion. The combination benefits from the blur invariance properties of the MLPQ-TOP descriptor. The fused system has been demonstrated to ourperform the state of the art face spoofing detection systems in most of the benchmarking tests adopted by the research community. Especially, on the CASIA FASD dataset containing video sequences that are relatively short or of low quality (a common case in practical scenarios) the fusion consistently improved the system performance. On the Replay-Attack and NUAA data sets, the fusion always led to better performance.

Future work will include a further investigation into the relationship between image quality and performance in spoofing detection. It would also be interesting to investigate in depth why MLPQ and, in particular, MBSIF, seem to work better than other dynamic texture descriptors. To gain this understanding, it will be necessary to develop effective methods for spatial and temporal registration, texture comparison and spatio-temporal visualisation.

Another future research direction relates to training. The ability of the BSIF representation (trained on completely general dynamic texture videos) to perform well in antispoofing applications is a very attractive attribute of the method. However, in future studies, it would be interesting to assess the merit of training SBIF filters using also spoofing attack data.

ACKNOWLEDGMENT

Partial support from the European Union project Beat is gratefully acknowledged.

REFERENCES

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

[2] B. Biggio, Z. Akhtar, G. Fumera, G. L. Marcialis, and F. Roli, "Security evaluation of biometric authentication systems under real spoofing attacks," *IET Biometrics*, vol. 1, pp. 11–24, 2012.

[3] Z. Akhtar, B. Biggio, G. Fumera, and G. L. Marcialis, "Robustness of multi-modal biometric systems under realistic spoof attacks against all traits," in *IEEE Workshop on Biometric Measurements and Systems for Security and Medical Applications (BioMS)*, Milan, Italy, 28/09/2011 2011, pp. 5–10.

[4] G. Chetty, "Biometric liveness detection based on cross modal fusion." in *FUSION*. IEEE, 2009, pp. 2255–2262.

[5] T. Kinnunen, Z. Wu, K.-A. Lee, F. Sedlak, E. Chng, and H. Li, "Vulnerability of speaker verification systems against voice conversion spoofing attacks: The case of telephone speech." in *ICASSP*. IEEE, 2012, pp. 4401–4404.

[6] N. Erdogmus and S. Marcel, "Spoofing face recognition with 3d masks," *Information Forensics and Security, IEEE Transactions on*, vol. 9, no. 7, pp. 1084–1097, July 2014.

[7] M. M. Chakka, A. Anjos, S. Marcel, R. Tronci, D. Muntoni, G. Fadda, M. Pili, N. Sirena, G. Murgia, M. Ristori, F. Roli, J. Yan, D. Yi, Z. Lei, Z. Zhang, S. Z. Li, W. R. Schwartz, A. Rocha, H. Pedrini, J. Lorenzo-Navarro, M. C. Santana, J. Mtt, A. Hadid, and M. Pietikinen, "Competition on counter measures to 2-d facial spoofing attacks." in *IJCB*, A. K. Jain, A. Ross, S. Prabhakar, and J. Kim, Eds. IEEE, 2011, pp. 1–6.

[8] I. Chingovska, J. Yang, Z. Lei, D. Yi, S. Li, O. Kahm, C. Glaser, N. Damer, A. Kuijper, A. Nouak, J. Komulainen, T. Pereira, S. Gupta, S. Khandelwal, S. Bansal, A. Rai, T. Krishna, D. Goyal, M.-A. Waris, H. Zhang, I. Ahmad, S. Kiranyaz, M. Gabbouj, R. Tronci, M. Pili, N. Sirena, F. Roli, J. Galbally, J. Ficrrcz, A. Pinto, H. Pedrini, W. Schwartz, A. Rocha, A. Anjos, and S. Marcel, "The 2nd competition on counter measures to 2d face spoofing attacks," in *Biometrics (ICB), 2013 International Conference on*, June 2013, pp. 1–6.

[9] T. de Freitas Pereira, J. Komulainen, A. Anjos, J. M. De Martino, A. Hadid, M. Pietikainen, and S. Marcel, "Face liveness detection using dynamic texture," *EURASIP Journal on Image and Video Processing*, vol. 2014:2, Jan. 2014.

[10] J. Kannala and E. Rahtu, "Bsif: Binarized statistical image features." in *ICPR*. IEEE, 2012, pp. 1363–1366.

[11] S. Arashloo and J. Kittler, "Class-specific kernel fusion of multiple descriptors for face verification using multiscale binarised statistical image features," *Information Forensics and Security, IEEE Transactions on*, vol. 9, no. 12, pp. 2100–2109, Dec 2014.

[12] ——, "Dynamic texture recognition using multiscale binarized statistical image features," *Multimedia, IEEE Transactions on*, vol. 16, no. 8, pp. 2099–2109, Dec 2014.

[13] L. Ghiani, A. Hadid, G. Marcialis, and F. Roli, "Fingerprint liveness detection using binarized statistical image features," in *Biometrics: Theory, Applications and Systems (BTAS), 2013 IEEE Sixth International Conference on*, Sept 2013, pp. 1–6.

[14] D. Cai, X. He, and J. Han, "Speed up kernel discriminant analysis," *The VLDB Journal*, vol. 20, no. 1, pp. 21–33, Feb. 2011.
[15] V. Ojansivu and J. Heikkil, "Blur insensitive texture classification using local phase quantization," in *Image and Signal Processing*, ser. Lecture Notes in Computer Science, A. Elmoataz, O. Lezoray, F. Nouboud, and D. Mammass, Eds. Springer Berlin Heidelberg, 2008, vol. 5099, pp. 236–243.
[16] Q. Zhen, D. Huang, Y. Wang, and L. Chen, "Lpq based static and dynamic modeling of facial expressions in 3d videos," in *Biometric Recognition*, ser. Lecture Notes in Computer Science, Z. Sun, S. Shan, G. Yang, J. Zhou, Y. Wang, and Y. Yin, Eds. Springer International Publishing, 2013, vol. 8232, pp. 122–129.
[17] B. Jiang, M. Valstar, B. Martinez, and M. Pantic, "A dynamic appearance descriptor approach to facial actions temporal modeling," *Cybernetics, IEEE Transactions on*, vol. 44, no. 2, pp. 161–174, Feb 2014.
[18] Z. Zhang, J. Yan, S. Liu, Z. Lei, D. Yi, and S. Z. Li, "A face antispoofing database with diverse attacks." in *ICB*, A. K. Jain, A. Ross, S. Prabhakar, and J. Kim, Eds. IEEE, 2012, pp. 26–31.
[19] 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 (6)*, ser. Lecture Notes in Computer Science, K. Daniilidis, P. Maragos, and N. Paragios, Eds., vol. 6316. Springer, 2010, pp. 504–517.
[20] G. Zhao and M. Pietikinen, "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, 2007.
[21] G. Pan, L. Sun, Z. Wu, and S. Lao, "Eyeblink-based anti-spoofing in face recognition from a generic webcamera." in *ICCV*. IEEE, 2007, pp. 1–8.
[22] "Zju eyeblink database," accessed: 2015-05-02. [Online]. Available: http://www.cs.zju.edu.cn/~gpan/database/db_blink.html
[23] 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. 3-4, pp. 215–225, 2011.
[24] S. Tirunagari, N. Poh, D. Windridge, A. Iorliam, N. Suki, and A. Ho, "Detection of face spoofing using visual dynamics," *Information Forensics and Security, IEEE Transactions on*, vol. 10, no. 4, pp. 762–777, April 2015.
[25] S. Bharadwaj, T. I. Dhamecha, M. Vatsa, and R. Singh, "Computationally efficient face spoofing detection with motion magnification," in *The IEEE Conference on Computer Vision and Pattern Recognition (CVPR) Workshops*, June 2013.
[26] M. D. Marsico, M. Nappi, D. Riccio, and J.-L. Dugelay, "Moving face spoofing detection via 3d projective invariants." in *ICB*, A. K. Jain, A. Ross, S. Prabhakar, and J. Kim, Eds. IEEE, 2012, pp. 73–78.
[27] K.-C. Lee, J. Ho, M.-H. Yang, and D. J. Kriegman, "Visual tracking and recognition using probabilistic appearance manifolds." *Computer Vision and Image Understanding*, vol. 99, no. 3, pp. 303–331, 2005.
[28] W. Bao, H. Li, N. Li, and W. Jiang, "A liveness detection method for face recognition based on optical flow field," in *Image Analysis and Signal Processing, 2009. IASP 2009. International Conference on*, April 2009, pp. 233–236.
[29] A. Anjos, M. M. Chakka, and S. Marcel, "Motion-based countermeasures to photo attacks in face recognition," *Institution of Engineering and Technology Journal on Biometrics*, Jul. 2013.
[30] Z. Zhang, D. Yi, Z. Lei, and S. Z. Li, "Face liveness detection by learning multispectral reflectance distributions." in *FG*. IEEE, 2011, pp. 436–441.
[31] J. Li, Y. Wang, T. Tan, and A. K. Jain, "Live face detection based on the analysis of fourier spectra," in *SPIE: Biometric Technology for Human Identification*, 2004, pp. 296–303.
[32] J. Mtt, A. Hadid, and M. Pietikinen, "Face spoofing detection from single images using micro-texture analysis." in *IJCB*, A. K. Jain, A. Ross, S. Prabhakar, and J. Kim, Eds. IEEE, 2011, pp. 1–7.
[33] 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 *ICB*, A. K. Jain, A. Ross, S. Prabhakar, and J. Kim, Eds. IEEE, 2012, pp. 67–72.
[34] W. R. Schwartz, A. Rocha, and H. Pedrini, "Face spoofing detection through partial least squares and low-level descriptors." in *IJCB*, A. K. Jain, A. Ross, S. Prabhakar, and J. Kim, Eds. IEEE, 2011, pp. 1–8.
[35] J. Komulainen, A. Hadid, M. Pietikainen, A. Anjos, and S. Marcel, "Complementary countermeasures for detecting scenic face spoofing attacks," in *International Conference on Biometrics*, Jun. 2013.
[36] D. Smith, A. Wiliem, and B. Lovell, "Face recognition on consumer devices: Reflections on replay attacks," *Information Forensics and Security, IEEE Transactions on*, vol. 10, no. 4, pp. 736–745, April 2015.
[37] D. Wen, H. Han, and A. Jain, "Face spoof detection with image distortion analysis," *Information Forensics and Security, IEEE Transactions on*, vol. 10, no. 4, pp. 746–761, April 2015.
[38] D. Menotti, G. Chiachia, A. Pinto, W. Robson Schwartz, H. Pedrini, A. Xavier Falcao, and A. Rocha, "Deep representations for iris, face, and fingerprint spoofing detection," *Information Forensics and Security, IEEE Transactions on*, vol. 10, no. 4, pp. 864–879, April 2015.
[39] J. Yang, Z. Lei, D. Yi, and S. Li, "Person-specific face antispoofing with subject domain adaptation," *Information Forensics and Security, IEEE Transactions on*, vol. 10, no. 4, pp. 797–809, April 2015.
[40] I. Chingovska and A. Rabello dos Anjos, "On the use of client identity information for face antispoofing," *Information Forensics and Security, IEEE Transactions on*, vol. 10, no. 4, pp. 787–796, April 2015.
[41] "Fast ica package," accessed: 2015-05-02. [Online]. Available: http://research.ics.aalto.fi/ica/fastica/
[42] G. Baudat and F. Anouar, "Generalized discriminant analysis using a kernel approach." *Neural Computation*, vol. 12, no. 10, pp. 2385–2404, 2000.
[43] B. Scholkopf and A. J. Smola, *Learning with Kernels: Support Vector Machines, Regularization, Optimization, and Beyond*. Cambridge, MA, USA: MIT Press, 2001.
[44] J. Zhang, M. Marszałek, S. Lazebnik, and C. Schmid, "Local features and kernels for classification of texture and object categories: a comprehensive study," *International Journal of Computer Vision*, vol. 73, no. 2, pp. 213–238, jun 2007.
[45] C. H. Chan, M. Tahir, J. Kittler, and M. Pietikainen, "Multiscale local phase quantization for robust component-based face recognition using kernel fusion of multiple descriptors," *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, vol. 35, no. 5, pp. 1164–1177, May 2013.
[46] M. Awais, F. Yan, K. Mikolajczyk, and J. Kittler, "Two-stage augmented kernel matrix for object recognition." in *MCS*, ser. Lecture Notes in Computer Science, C. Sansone, J. Kittler, and F. Roli, Eds., vol. 6713. Springer, 2011, pp. 137–146.
[47] F. Yan, J. Kittler, K. Mikolajczyk, and M. A. Tahir, "Non-sparse multiple kernel fisher discriminant analysis." *Journal of Machine Learning Research*, vol. 13, pp. 607–642, 2012.
[48] A. F. Martin, G. R. Doddington, T. Kamm, M. Ordowski, and M. A. Przybocki, "The det curve in assessment of detection task performance." in *EUROSPEECH*, G. Kokkinakis, N. Fakotakis, and E. Dermatas, Eds. ISCA, 1997.
[49] J. Galbally and S. Marcel, "Face anti-spoofing based on general image quality assessment," in *Proc. IAPR/IEEE Int. Conf. on Pattern Recognition, ICPR*, August 2014.
[50] T. de Freitas Pereira, A. Anjos, J. M. D. Martino, and S. Marcel, "Lbp - top based countermeasure against face spoofing attacks." in *ACCV Workshops (1)*, ser. Lecture Notes in Computer Science, J.-I. Park and J. Kim, Eds., vol. 7728. Springer, 2012, pp. 121–132.
[51] Y. Jianwei, L. Zhen, L. Shengcai, and S. Z. Li, "Face liveness detection with component dependent descriptor," in *Proceedings of the 6th IAPR International Conference on Biometrics, (ICB)*, 2013.
[52] T. de Freitas Pereira, A. Anjos, J. De Martino, and S. Marcel, "Can face anti-spoofing countermeasures work in a real world scenario?" in *Biometrics (ICB), 2013 International Conference on*, June 2013, pp. 1–8.

Shervin Rahimzadeh Arashloo obtained his Ph.D. in computer vision from the Centre for Vision, Speech and Signal Processing (CVSSP) at university of Surrey. His research interests include biometrics, graphical models and cognitive vision. He is currently an assistant professor of electrical engineering with Umria university and a visiting senior fellow with university of Surrey.

Josef Kittler is Professor of Machine Intelligence at 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 a Prentice Hall textbook on Pattern Recognition: A Statistical Approach, as well as more than 170 journal papers. He serves on the Editorial Board of several scientific journals in Pattern Recognition and Computer Vision.