Unseen Face Presentation Attack Detection Using Sparse Multiple Kernel Fisher Null-Space

Shervin Rahimzadeh Arashloo

***Abstract*— We address the face presentation attack detection problem in the challenging conditions of an unseen attack scenario where the system is exposed to novel presentation attacks that were not available in the training stage. To this aim, we pose the unseen face presentation attack detection (PAD) problem as the one-class kernel Fisher null-space regression and present a new face PAD approach that only uses bona fide (genuine) samples for training. Drawing on the proposed kernel Fisher null-space face PAD method and motivated by the availability of multiple information sources, next, we propose a multiple kernel fusion anomaly detection approach to combine the complementary information provided by different views of the problem for improved detection performance. And the last but not the least, we introduce a sparse variant of our multiple kernel Fisher null-space face PAD approach to improve inference speed at the operational phase without compromising much on the detection performance. The results of an experimental evaluation on the OULU-NPU, Replay-Mobile, Replay-Attack and MSU-MFSD datasets illustrate that the proposed method outperforms other methods operating in an unseen attack detection scenario while achieving very competitive performance to multi-class methods (that benefit from presentation attack data for training) despite using only bona fide samples in the training stage.**

***Index Terms*— Face presentation attack detection, anti-spoofing, unseen attacks, novelty detection, one-class classification, kernel regression, multiple kernel fusion, sparse regularisation.**

I. Introduction

THE face recognition technology has made great advances during the past couple of decades [1]–[4]. Nevertheless, the functionality of these systems in practical situations is compromised by their susceptibility to presentation attacks (PA's) where an imposter tries to be authenticated as a genuine client by presenting fake biometric traits to system sensors. Due to potential security risks associated with the problem, face presentation attack detection (a.k.a. anti-spoofing) has received increasing attention over the past years, resulting in a variety of different countermeasures [5], [6]. A majority of the existing approaches assume that the face presentation attack detection (PAD) problem is a closed-set recognition task, and subsequently formulate and train a binary classifier using the available positive (bona fide) and negative (PA) training samples. Nevertheless, even with the impressive performances reported on some databases, the technology is not mature yet. The discrepancies in samples due to varying image acquisition settings, including different environmental conditions, sensor interoperability issues, etc. degrade the performance of face PAD techniques. In this context, a particularly challenging facet of the problem is due to the previously "unseen" attack types for which no similar training samples in terms of the presentation attack instruments are available at training time. In these situations, the common closed-set two-class formulation of the problem tends to generalise poorly, degrading the performance of face PAD methods. The unseen PA detection problem is not only pertinent to the face modality [7]–[12] but also to other biometric modalities including fingerprint [13]–[16], iris [17], voice [18], etc. From a general perspective beyond biometrics, different attacks including adversarial attacks may pose serious challenges to the operation of different classification systems including deep learning methods [19].

While early face PAD benchmark datasets only included a single attack type (typically printed paper), more recent databases incorporate a more diverse set of attacks, including digital photo attacks, replay attacks, mask attacks, make-up attacks, etc. Although introducing new datasets [20], [21] that cover a wider variety of possible presentation attack mechanisms and instruments is desirable, yet, it may not completely solve the problem. The limitation arises from the fact that not all possible attack scenarios can be anticipated and covered in the datasets since there is always the possibility of developing a new attack strategy to bypass the existing countermeasures. Consequently, the error rates obtained on one or more datasets under a closed-set attack assumption may not be generalised and regarded as representative error rates corresponding to a real-world operation of the system.

In practice, presentation attacks may potentially appear as fairly diverse, or novel and unpredictable while bona fide sample distributions tend to have relatively less diversity. Notwithstanding other strategies, an alternative approach to the face PAD problem is to try to capture the distribution of bona fide samples. This objective may be realised through a one-class classification (OCC) [22] formalism to identify patterns from a target class, conforming to a specific condition, and differentiate them from non-target objects. OCC differs from the conventional multi-class formulation in that it mainly relies on observations from a single (target) class for training. In the context of biometric PAD, this approach has been examined,

Manuscript received September 12, 2020; revised November 21, 2020; accepted December 16, 2020. Date of publication December 22, 2020; date of current version October 4, 2021. This article was recommended by Associate Editor X. Wang.

The author is with the Department of Computer Engineering, Faculty of Engineering, Bilkent University, 06800 Ankara, Turkey (e-mail: s.rahimzadeh@cs.bilkent.edu.tr).

Color versions of one or more figures in this article are available at https://doi.org/10.1109/TCSVT.2020.3046505.

Digital Object Identifier 10.1109/TCSVT.2020.3046505

for instance, in speech [23], fingerprint [15], or iris [24] anti-spoofing as well as in face PAD [7], [8] by considering bona fide samples as target objects and then trying to detect PA's as novelties in reference to the population of target observations.

Motivated by the merits of an OCC formulation in detecting unseen presentation attacks, this study follows a one-class classification methodology to the face PAD problem. For this purpose, we approach the problem in a reproducing kernel Hilbert space (RKHS) and formulate the unseen face PAD as a one-class kernel Fisher null-space OCC problem. By virtue of achieving an optimal Fisher classification criterion (i.e. zero within-class and positive between-class scatters) and in spite of operating in a pure one-class approach that avoids the use of negative training samples altogether, the proposed method achieves a superior detection performance as compared with other OCC face PAD methods in the literature. Operating in a pure one-class scenario and circumventing the use of negative training samples is particularly important to assess the generalisation capability of the method in practical real-world settings.

While ensemble techniques have been widely deployed in the literature to enhance the classification performance in different problem domains [25], their application to the face PAD problem has been very limited. In this context, formulation of the unseen face PAD problem as a kernel-based method in this work opens the door to benefit from multiple complementary sources of information for improved performance. In this respect, we propose a multiple kernel fusion mechanism to combine different views of the problem at hand each of which is derived by introducing diversity in the representations obtained from face images. While a kernel fusion approach for face PAD has not been considered previously and is novel, its efficacy in improving the performance of the proposed Fisher null-space face PAD approach (as well as other widely used one-class classifiers) is verified thoroughly on different datasets in this work.

A drawback of kernel-based approaches is that the computational complexity of these methods in the operational phase which scales linearly in the number of training samples. As a final contribution, we advocate a sparse variant of the proposed face PAD approach and illustrates its utility in reducing the complexity of the naïve approach by tens of orders of magnitude without compromising much on the detection performance.

In summary, the main contributions of the present study are as follows:

1) We formulate the unseen face PAD problem as a one-class kernel Fisher null-space regression anomaly detection problem. Despite operating in a pure one-class paradigm and avoiding the use of negative training samples altogether, the proposed formulation is shown to be superior to the existing one-class unseen face PAD methods.
2) We propose a multiple kernel fusion anomaly detection strategy to combine the complementary information provided by different views of the problem for improved detection performance. The improvements obtained by the proposed kernel fusion approach underlines the importance of fusing multiple information sources in an OCC face PAD approach.
3) We present a sparse representation-based variant of the proposed multiple kernel fusion one-class Fisher null-space approach to speed up inference at the operational stage.
4) Finally, a thorough evaluation of the proposed face PAD approach is carried out on four publicly available datasets in an unseen (zero-shot) presentation attack detection scenario demonstrating the competitive performance of the proposed approach against other one-class as well as multi-class face PAD techniques.

The rest of the paper is organised as follows: Section II reviews related work with an emphasis on the unseen face PAD approaches. In Section III, the proposed one-class approach for unseen face PAD based on kernel regression is introduced where multiple kernel fusion, sparse regularisation of the solution and other related technical aspects of approach are discussed. The results corresponding to an experimental assessment of the proposed method are presented in Section V. Finally, conclusions are drawn in Section VI.

II. Related Work

Different countermeasures including those based on special-purpose hardware, software and challenge-response methods have been proposed for face presentation attack detection [26]. The software-based approaches classify an image (sequence) based on different features derived from image content. In this study, we follow a software-based approach to face PAD using a single modality (i.e. visible spectrum RGB images) for PA detection in contrast to some other studies operating beyond the visible spectrum [27]. Texture is the most commonly deployed cue for face PAD among the features derived from an image/image sequence [28], [29]. A different category of face PAD methods constitutes motion-based approaches [30]–[32]. An alternative category of methods relates to frequency-based approaches to detect PA's in the Fourier domain [32]–[34]. Colour characteristics [35], [36] as well as shape information are also utilised as different mechanisms to detect presentation attacks [37]. As bona fide and PA attempts appear differently under the same illumination conditions, some other methods [38] use reflectance for face PAD. Other work [39] uses a statistical model of image noise for face PAD. A different study [40] develops a generic classifier for detecting face presentation attack images by focusing on the contours of the spoofing medium. For this purpose, face presentation attack detection is posed as the problem of detecting contours in the image. The authors then train a deep CNN network to classify an image as a bona fide or attack sample by measuring the probability of incorporating spoofing contours in an object. The results of an experimental evaluation on two datasets confirms the effectiveness of the approach in cross-database scenarios. Nevertheless, as the method relies on spoofing medium contours, its applicability is limited to certain cases where such contours are visible in the image. A recent category of approaches relates to deep learning

methods, and in particular, deep convolutional neural networks (CNN's) [41]–[43].

In terms of the two-class classification schemes, different alternatives have been examined. These include discriminant classifiers including the Support Vector Machines [44], [45], the linear discriminant analysis [35], [46], neural networks [31], convolutional neural networks [41], [42], Bayesian networks [47] as well as Adaboost [48]. An alternative category includes regression-based approaches trying to project input features onto their labels [49]. Methods trying to learn a distance metric [50] as well as some heuristic methods have been also examined for classification in face PAD [30], [51].

In contrary to the common close-set two-class formulation of the problem, there also exists a different category of approaches trying to address the face PAD problem in an unseen attack scenario. One main group of these methods formulates the problem as an OCC task to detect unseen PA's [7]. The work in [8] considers a Gaussian mixture model (GMM) one-class learner for classification operating on image quality measures. A different study [9] analyses One-Class SVM and Auto Encoder-based classifiers to address unseen face PAD. The work in [10] examines a new strategy to utilise the client-specific information to train one-class classifiers using only bona fide data. In an alternative study [11], detection of unknown PA's was addressed in a Zero-Shot Face Anti-spoofing (ZSFA) scenario via a deep tree network (DTN) to partition the PA samples into semantic sub-groups in an unsupervised fashion. The work in [12] introduces a method where a deep metric learning model is proposed using a triplet focal loss as regularisation for the so-called metric-softmax approach.

For a more detailed review of the face PAD methods one may consult [5], [6], [52].

III. Kernel Regression for One-Class Face PAD

In this section, first, a brief background on regression in the reproducing kernel Hilbert space (kernel regression) shall be provided. Next, kernel regression is used as a tool for face PA anomaly detection.

Let $\mathcal{F}$ be a feature space induced by a non-linear mapping $\boldsymbol{\phi} : \mathbb{R}^d \rightarrow \mathcal{F}$. For a suitably chosen mapping, an inner product $\langle ., . \rangle$ on $\mathcal{F}$ may be represented as $\langle \boldsymbol{\phi}(x_i), \boldsymbol{\phi}(x_j) \rangle = \kappa(x_i, x_j)$, where $\kappa(., .)$ is a positive semi-definite kernel function. In kernel regression, each point x is first projected onto $\phi(x)$ followed by seeking a real-valued function $g(\phi(x)) = f(x)$ minimising a sum of squared differences between the expected and the generated responses. The relation for $f(z)$ may be written as

$$\begin{aligned} f(z) &= [\langle \phi(z), \phi(x_1) \rangle, \ldots, \langle \phi(z), \phi(x_n) \rangle](\phi(\mathbf{X})\phi(\mathbf{X})^\top)^{-1}\mathbf{y} \\ &= [\kappa(z, x_1), \ldots \kappa(z, x_n)]\mathbf{K}^{-1}\mathbf{y} \end{aligned} \tag{1}$$

where we have used the notation $\mathbf{K} = \phi(\mathbf{X})\phi(\mathbf{X})^\top$ to denote the so-called kernel matrix. Denoting $\boldsymbol{\alpha} = \mathbf{K}^{-1}\mathbf{y}$, function $f(.)$ may be represented as

$$f(.) = [\kappa(., x_1), \ldots \kappa(., x_n)]\boldsymbol{\alpha} = \sum_{i=1}^{n} \alpha_i \kappa(., x_i). \tag{2}$$

The full responses on the training set $\mathbf{X}$ may be derived as $f(\mathbf{X}) = \mathbf{K}\boldsymbol{\alpha}$ and the corresponding cost function in this case is $\|\mathbf{K}\boldsymbol{\alpha} - \mathbf{y}\|_2^2$ where $\|.\|_2^2$ denotes the squared l_2-norm.

In a one-class classification task it is desirable to have normal samples forming a compact cluster while being distant from anomalies. In a reproducing kernel Hilbert space (RKHS) and in the absence of outlier training data, in a one-class classification paradigm, it is common practice to consider the origin as an artificial exemplar outlier with respect to the distribution of positive samples [53].

In this work, similar to [54], a projection function (defined in terms of kernel regression) is used such that it maps bona fide samples onto a compact cluster, distant from a hypothetical non-target observation lying at the origin. This objective may be achieved by setting the responses for all target observations to a common fixed real number, distinct from zero, i.e. $y_i = c$, for all i, s.t. $c \neq 0$. In this case, the kernel regression approach would form a compact cluster of target samples as they would be all mapped onto the same point. Note that kernel regression performs an exact interpolation when the parameters characterising the regression (i.e. $\boldsymbol{\alpha}$) can be uniquely determined. This is the case when the kernel matrix is positive-definite. Since by assumption $c \neq 0$, the projected normal observations would lie away from the (hypothetical) outlier, i.e. the origin. Without loss of generality one may set $c = 1$, as the exact value for c would only act as a scaling factor. Having set the response vector $\mathbf{y}$ to $\mathbf{1}^{n\times 1}$, one may solve for $\boldsymbol{\alpha}$ as $\boldsymbol{\alpha} = \mathbf{K}^{-1}\mathbf{y} = \mathbf{K}^{-1}\mathbf{1}^{n\times 1}$.

The procedure discussed above provides the best separability of normal samples from outliers with respect to the Fisher criterion [54], stated formally in the following proposition.

Proposition: Assuming the origin as a hypothetical outlier, the kernel regression approach with the responses for all target samples set to a common fixed real number other than zero, corresponds to the optimal kernel Fisher criterion for classification (i.e. kernel Fisher null-space), resulting in a zero within-class variance while providing a positive between-class scatter.

Proof: In a Fisher classifier, the objective function may be expressed as the ratio of the inter-class scatter to the total intra-class scatter as [55]

$$J = \frac{s_B}{s_W} = \frac{(m_1 - m_2)^2}{s_1^2 + s_2^2} \tag{3}$$

where m_1 and m_2 denote the mean of the first and the second class, respectively while s_1^2 and s_2^2 represent the corresponding within-class scatters. As noted earlier, the Fisher analysis, originally developed for a binary classification problem, may be applied to the one-class scenario by assuming the origin as an exemplar outlier. As all positive samples are projected onto the same point (i.e. 1), the associated mean of the positive class would be $m_1 = 1$. In this case, the within-class scatter of the transformed bona fide samples is

$$s_1^2 = \sum_{C_1} (y_i - m_1)^2 = \sum_{C_1} (1 - 1)^2 = 0 \tag{4}$$

where C_1 stands for the bona fide class while y_i denotes the mapping of the observation $x_i \in C_1$ onto the subspace.

Regarding the PA class, only a single hypothetical sample is assumed to exist at the origin and hence the average of the negative class shall be $m_2 = 0$ while its within-class scatter is $s_2 = 0$. The total within-class scatter in this case would then be $s_W = s_1^2 + s_2^2 = 0$.

Regarding the between-class scatter (the numerator of J in Eq. 3) we have

$$s_B = (m_1 - m_2)^2 = (1 - 0)^2 = 1. \tag{5}$$

Thus, as discussed in [54], the one-class kernel regression when the responses of all target (bona fide) samples are equal and distinct from zero, corresponds to a projection function (i.e. $f(.)$) leading to $s_W = 0$ and $s_B = 1$, and consequently, represents a kernel Fisher null-space analysis [56], [57]. ■

The foregoing kernel regression-based classification approach may be considered as a *one-class* variant of our earlier two-class kernel Fisher analysis [58] introduced for face PAD and that of [59] developed for face verification.

A. Regularisation

Formulating a one-class classifier based on the kernel regression formalism opens the possibility to regularise the solution. Regularising the solution of a regression problem may be driven by different objectives. First, when the number of observations is smaller than the number of variables, the least-squares problem is ill-posed. In this case, additional constraints are introduced into the problem to uniquely specify the solution. The second scenario is when the model suffers from poor generalisation. In this case, regularisation improves the generalisation capability of the model by introducing a limitation on the available function space via imposing a penalty to discourage specific regions of the function space. This latter case corresponds to imposing priors on the solution to maintain a desired trade-off between data fidelity and some condition on the solution.

A regularised kernel regression problem may be expressed as finding the vector minimser of the following cost function:

$$Q(\boldsymbol{\alpha}) = \|\mathbf{K}\boldsymbol{\alpha} - \mathbf{y}\|_2^2 + \mathcal{R}(\boldsymbol{\alpha}) \tag{6}$$

where, as before, $\|\mathbf{K}\boldsymbol{\alpha} - \mathbf{y}\|_2^2$ measures the closeness of the generated responses to the expected responses $\mathbf{y}$ while $\mathcal{R}(\boldsymbol{\alpha})$ encodes a desired regularisation on the solution $\boldsymbol{\alpha}$.

Regularisation schemes favouring sparseness of the solution are among the widely applied techniques [60]. These methods try to represent an observation in terms of a few atoms from a given dictionary. Sparsity of the solution to a least squares problem may be encouraged via an l_1-norm, which, in statistics, is known as Lasso and in the signal processing community is referred to as a basis pursuit. In addition to enhanced generalisation performance, sparse L_1-norm models provide scalable techniques that can be used in large-scale problems. This is particularly important in kernel-based methods as the complexity of these techniques grows linearly in the number of training samples. Thus, besides other benefits discussed above, a sparse solution is also advantageous in enhancing the computational complexity of the method. In this work, sparseness of the solution of the one-class kernel regression is encouraged by prescribing an L_1-regulariser on $\boldsymbol{\alpha}$, i.e. $\mathcal{R}(\boldsymbol{\alpha}) = \delta \sum_{i=1}^{n} |\alpha_i|$. The objective function for the sparse one-class kernel regression may then be expressed as

$$Q(\boldsymbol{\alpha}) = \|\mathbf{K}\boldsymbol{\alpha} - \mathbf{y}\|_2^2 + \delta \sum_{i=1}^{n} |\alpha_i| \tag{7}$$

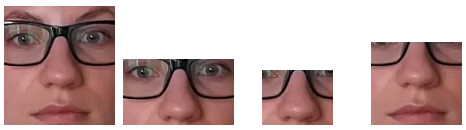
where parameter δ controls the sparseness of the solution. By imposing a strong sparseness prior on $\boldsymbol{\alpha}$, each response y_i would be characterised by only a few samples from among the training set. A desirable outcome of a sparse representation is a reduction in the computational complexity in the operational stage, determined by the number of non-zero elements of $\boldsymbol{\alpha}$.

IV. One-Class Multiple Kernel Fusion

In kernel-based methods, the kernel function plays an important role as it specifies the embedding of the data in the feature space. While ideally the kernel function and consequently the corresponding embedding is to be learnt directly from training samples, in practice, a relaxed alternative of the problem is considered by trying to learn an optimal combination of multiple kernels providing different views of the problem at hand. The coefficients characterising the combination may then be learnt using training instances from multiple classes [61]. In a one-class novelty detection approach and in the absence of non-target (PA) training samples (i.e. unseen face PAD), we opt for the average fusion of multiple base kernels. In the proposed approach, diverse views of the face PAD problem are constructed following two mechanisms: using multiple (local) face image regions and deployment of multiple representations derived from these regions, discussed next.

Multiple Regions: In addition to the whole face image providing discriminatory information at a global level, different local regions of the face image convey distinctive information and characteristics for face PAD decision making. In order to benefit from local information, along with the whole face image tightly cropped to minimise the background effects (identified as region R1), three additional local regions are also considered. These include eyes and nose as region R2, nose and the surrounding as region R3 and region R4 which focuses on the areas surrounding the nose and the mouth, Fig. 1. In this study, the regional representations corresponding to main facial features are considered. One may also consider increasing local regions. However, simply increasing the number of regional representations does not guarantee improved performance as inclusion of less informative representations to the multiple kernel fusion approach in a uniformly weighted combination scheme followed in this study may disturb the fusion mechanism and may even deteriorate the performance.

Multiple Representation: The second source of diversity for multiple kernel fusion is derived through using different image representations. While a wide range of different image features including LBP and its variants [59], [62], image quality measures [8], [62], etc. are applied to face PAD, more recently, a great deal of research has been directed towards investigating the applicability of deep convolutional neural network (CNN) representations for face PAD [43]



(a) (b) (c) (d)

Fig. 1. Multiple regions used to derive representations: (a): region R1; (b): region R2; (c): region R3; (d): region R4.

illustrating that such features may provide discriminative information for detection of presentation attacks. The application of CNN-based representation may also simplify the overall design architecture of biometric systems through adoption of common representation for the face matching problem. Following the same methodology, the current study utilises CNN representations as features for face PAD. For this purpose, features obtained from the penultimate layers of pre-trained GoogleNet [63], ResNet50 [64] and VGG16 [65] networks are used to construct multiple representations for each facial region.

A. Learning the Discriminant

Once multiple representations from different regions are derived, the learning process of the proposed subject-specific sparse one-class multiple kernel fusion regression approach for subject c is given as

$$\boldsymbol{\alpha}^c = \arg\min_{\boldsymbol{\alpha}} \| \frac{1}{RN} \sum_{r=1}^{R} \sum_{n=1}^{N} \mathbf{K}_{rn}^c \boldsymbol{\alpha} - \mathbf{y} \|_2^2 + \delta \sum_{i=1}^{n} |\alpha_i| \quad (8)$$

where $R = 4$ and $N = 3$ denote the number of facial regions and deep CNN networks used for face image representation, respectively while $\mathbf{K}_{rn}^c$ stands for the kernel matrix associated with region r whose representation is obtained via CNN with index n. In a subject-specific approach, the kernel matrices $\mathbf{K}_{rn}^c$'s are constructed by using the training instances (i.e. features extracted from bona fide frames) of subject c only. In this study, Eq. 8, is solved using the Least Angle Regression (LARS) algorithm [66] which facilitates deriving solutions with all possible cardinalities on $\boldsymbol{\alpha}$.

B. Decision Strategy

Once the projection parameter $\boldsymbol{\alpha}^c$ is inferred for client c, the projection of a test sample (z) onto the feature subspace of subject c is given as

$$f^c(z) = \frac{1}{RN} \sum_{i=1}^{M} \alpha_i^c \sum_{r=1}^{R} \sum_{n=1}^{N} \kappa_{rn}(z, \mathbf{x}_i^c) \quad (9)$$

where α_i^c denotes the i^{th} (non-zero) element of the discriminant in the Hilbert space for the c^{th} subject while $\mathbf{x}_i^c$ denotes the i^{th} training instance of subject c. M is the total number of non-zero elements of $\boldsymbol{\alpha}$. $\kappa_{rn}(z, \mathbf{x}_i^c)$ is the kernel function, capturing the similarity between the r^{th} region of the test sample z and that of the i^{th} training instance of subject c based on the representation derived through the n^{th} deep CNN network.

1) Raw Score Fusion: Eq. 9 provides the raw projection score for a single frame of a test video sequence. In order to derive a score for the whole test video, one possibility is to simply average the raw scores corresponding to individual frames comprising the video, leading to a decision rule as

$$\begin{aligned} \frac{1}{F} \sum_{f=1}^{F} f^c(z_f) &\geq \tau^c \qquad \textit{bona fide,} \\ \frac{1}{F} \sum_{f=1}^{F} f^c(z_f) &< \tau^c \qquad \textit{PA} \end{aligned} \quad (10)$$

where F denotes the total number of frames in a video sequence while τ^c is the threshold for decision making for subject c.

2) Fusion of Probabilistic Scores: Next, we approximate the probability density function of the score distributions corresponding to bona fide samples using a Gaussian distribution and use the cumulative density function of the inferred Gaussian as a probabilistic measure of normality. In this case, the decision rule for a video sequence is defined as

$$\begin{aligned} \frac{1}{F} \sum_{f=1}^{F} \int_{-\infty}^{z_f} \mathcal{N}_{\mu,\sigma}(x)\, dx &\geq \tau^c \qquad \textit{bona fide,} \\ \frac{1}{F} \sum_{f=1}^{F} \int_{-\infty}^{z_f} \mathcal{N}_{\mu,\sigma}(x)\, dx &< \tau^c \qquad \textit{PA} \end{aligned} \quad (11)$$

where $\mathcal{N}_{\mu,\sigma}$ denotes a normal distribution with mean μ and standard deviation σ.

V. Experimental Evaluation

In this section, the results of an experimental evaluation of the proposed approach on four publicly available datasets in an "unseen" attack scenario are presented. Driven by the success of client-specific modelling for face PAD [10], [67], we build a separate model for each individual client in the dataset.

A. Implementation Details

A number of details regarding implementation of the proposed approach are in order. Each frame is initially preprocessed using the photometric normalisation method of [68] to compensate for illumination variations. The face detection bounding boxes provided along with each dataset are used to locate the face in each frame. The coordinates associated with the last detected face in a video sequence is used for a frame missing a bounding box. In order to select facial regions (Fig. 1) in a consistent fashion across different frames, the OpenFace library [69] is used to detect landmarks around facial features. The deep CNN representations yield 1024-, 2048- and 4096-dimensional feature vectors for the GoogLenet (N1), ResNet50 (N2) and VGG16 (N3) networks, respectively. The thus obtained feature vectors are normalised to have a unit L_2-norm. The kernel function used is that of a Gaussian (i.e. $\kappa(x_i, z) = \exp(-\theta \|x_i - z\|_2^2)$ kernel yielding a positive definite kernel matrix where θ is set to the reciprocal of the average pairwise Euclidean distance between bona fide

training samples. The proposed method is implemented as unoptimised Matlab codes running on a 64-bit 4.00GHz Intel Core-i7 Ubuntu machine with 32GB memory.

B. Datasets

The datasets used in the current work are briefly introduced next.

1) The Replay-Mobile Dataset: The Replay-Mobile dataset [70] includes 1190 video recordings of both bona fide and attack samples of 40 individuals recorded under different illumination conditions using two different acquisition devices. The dataset is divided into three disjoint subsets of training, development and testing and an additional partition corresponding to enrolment data.

2) The Replay-Attack Dataset: The Replay-Attack database [71] includes 1300 video recordings of bona fide and attack samples of 50 different individuals. Attacks are created either using a printed image, a mobile phone or a high definition iPad screen. The data is randomly divided into three subsets for training (60 bona fide and 300 PA samples), development (60 bona fide and 300 PA samples) and testing (80 bona fide and 400 PA samples) purposes without any overlap between subjects of different sets.

3) The OULU-NPU Dataset: The OULU-NPU database [72] consists of 4950 bona fide and attack video recordings of 55 subjects recorded using six different cameras in three sessions with different illumination conditions and backgrounds. The dataset includes previously unseen input sensors, attack types and acquisition conditions. The videos of the 55 subjects in this database are divided into three subject-disjoint subsets for training (360 bona fide and 1440 PA samples), development (270 bona fide and 1080 PA samples) and testing (360 bona fide and 1440 PA samples) purposes. For the evaluation purposes, four protocols are designed amongst which, the forth protocol is the most challenging one which is used in this study.

4) The MSU-MFSD Dataset: The MSU MFSD database [73] includes 440 videos of photo and video attack attempts of 55 individuals recorded using two different cameras. The publicly available subset, however, includes 35 subjects. The dataset is divided into two subject-disjoint sets for training (30 bona fide and 90 PA's) and testing (40 bona fide and 120 PA's) purposes.

C. Performance Metrics

For performance reporting ISO metrics BSISO-IEC30107-3-2017 are used (APCER↓, BPCER↓ and ACER↓). In addition to the ISO metrics, the efficacy of the proposed method is also reported in terms of AUC (the Area Under the ROC Curve ↑) as well as the Half Total Error Rate (HTER↓) and the Equal Error Rate (EER↓), whenever required for a comparison to other techniques.

D. Effect of Multiple Kernel Fusion

First, the effect of fusing multiple views of the problem via a kernel fusion strategy is analysed. For this purpose and in order to summarise the performance of each individual kernel to facilitate a comparison, ACER's corresponding to different facial regions and different CNN representations are reported in Table I where RxNy denotes the representation for region $x \in \{1, \ldots, 4\}$ derived from the y^{th} deep CNN (N_1 = GoogLenet, N_2 = ResNet50, N_3 = VGG16). From the table, the following observations may be made. First, the best performing regional representation among others in terms of average ACER is R4N3 (i.e. VGG16 applied to the region focusing on the areas surrounding the nose and the mouth). Interestingly, the performance of R4 is better than the whole face image (R1) using the same deep CNN representation. Second, among three different deep representations one observes that the VGG16 network provides the most discriminative features for presentation attack detection. Third, regardless of the region and the deep CNN features employed, the Replay-Mobile and the OULU-NPU (forth protocol) appear to be more challenging as compared with other datasets. In terms of the kernel fusion strategy, the average ACER for the fusion is 5.58% whereas the best performing single kernel provides an average ACER of 7.01% while the worst performing kernel yields an average ACER of 12.83%. That is, the kernel fusion improves the average ACER w.r.t. the best single kernel by more than 20% and by more than 56% w.r.t. the worst performing single kernel.

TABLE I

EFFECT OF MULTIPLE KERNEL FUSION IN TERMS OF ACER (%). R.M.:REPLAY-MOBILE; R.A.: REPLAY-ATTACK; M.M.: MSU-MFSD; O.N.: OULU-NPU; ALL: AVERAGE OVER ALL DATASETS

Kernel	R.M.	R.A.	M.M.	O.N.	All
R1N1	20.00	3.33	0.04	13.75	9.28
R2N1	15.23	3.12	0.04	10.42	7.20
R3N1	18.18	8.13	0.04	7.50	8.46
R4N1	18.41	8.12	0.04	9.17	8.93
R1N2	30.45	1.67	0.04	11.25	10.85
R2N2	25.68	0.21	0.04	25.42	12.83
R3N2	18.64	0.21	0.04	20.42	9.82
R4N2	19.09	0	0.04	12.08	7.80
R1N3	20.45	0.42	0.04	10.42	7.83
R2N3	24.55	0	0.04	8.33	8.23
R3N3	18.64	0.42	0.04	11.67	7.69
R4N3	19.90	0.63	0.04	7.50	7.01
Fusion	15.68	0	0	6.67	5.58

E. Effect of Sparse Regularisation

Next, the effectiveness of sparse regularisation on the one-class kernel regression approach is examined. For this purpose, using the LARS algorithm [66], solutions ($\boldsymbol{\alpha}$) with different cardinalities (number of non-zero elements) of 2, 3, 4, 5, 10, 20, 30 and 50 are obtained. Cardinalities of higher than 50 are found not to improve the overall average ACER. The performances corresponding to different cases in terms of ACER are reported in Table II. The last column of the table (ARSG), reports the Average Relative Speed-up Gain achieved compared to the non-sparse solution in the test phase. From the table, it may be observed that, interestingly, even by using 2 training frames, one may achieve an impressive overall average

TABLE II

EFFECT OF SPARSITY IN TERMS OF ACER (%). NNZ: NUMBER OF NON-ZERO ELEMENTS OF $\boldsymbol{\alpha}$; R.M.: REPLAY-MOBILE; R.A.: REPLAY-ATTACK; M.M.: MSU-MFSD; O.N.: OULU-NPU; ALL: AVERAGE OVER ALL DATASETS; ARSG: THE AVERAGE RELATIVE SPEED UP GAIN

NNZ	R.M.	R.A.	M.M.	O.N.	All	ARSG	Time(ms)
2	21.36	0.0	5.63	5.42	8.10	$\approx 390\times$	0.015
3	18.86	0.0	0.0	5.83	6.17	$\approx 260\times$	0.023
4	17.50	0.0	0.0	6.67	6.04	$\approx 200\times$	0.030
5	15.68	0.0	0.0	6.67	5.58	$\approx 160\times$	0.038
10	16.14	0.0	0.0	7.50	5.91	$\approx 80\times$	0.077
20	15.68	0.0	0.0	9.17	6.21	$\approx 40\times$	0.1547
30	15.23	0.0	0.0	9.17	6.10	$\approx 25\times$	0.2320
50	15.23	0.0	0.0	10.0	6.30	$\approx 15\times$	0.3868

ACER of 8.10%. Increasing the cardinality from 2 towards 5, a reduction in the average ACER over four datasets is achieved where by using only 5 bona fide training frames, the proposed approach yields an average ACER of 5.58%. Increasing the cardinality beyond 5 towards 50, no further improvement in terms of the overal average ACER is obtained. Regarding the ARSG, the best performing method in terms of average ACER with NNZ=5 (NNZ: Number of Non-Zero elements in $\boldsymbol{\alpha}$), yields an impressive $160\times$ speed-up gain in the test phase compared to its non-sparse counterpart. The inference time (in milliseconds), excluding the feature extraction step, for different NNZ's are reported in the rightmost column of Table II. Considering the average performance over four datasets and the inference times, the sparse solution with NNZ=5 seems to be good trade-off between computational complexity and performance. The inference time for the case of NNZ=5 is 0.038 milliseconds. Note that, however, the reported inference times are obtained on a PC with the specifications as described in Section V-A using conventional serial code. Since the decision strategy of the proposed approach is amenable to parallel processing (see Eq. 9), in practical situations, during the testing phase, the inference time may be substantially reduced by porting the corresponding computations onto a graphical processing unit (GPU) for parallel processing.

F. Effect of Temporal Aggregation

The impact of temporal aggregation of frame-level scores to derive a video-level decision using a sum fusion rule over raw scores and over probabilistic scores is analysed in this section. The frame-level and video-level performances of the proposed approach on different datasets along with the average ACER's are reported in Table III. From the table it may be observed that the sum fusion over raw scores (Video-level (raw)) improves the average ACER on four datasets from 8.99% to 5.58%. That is, more than 37% reduction in the average ACER. The average fusion rule applied to the probabilistic scores yields an even further reduction in the average ACER, from 5.58% to 4.97%, corresponding to more than 10% improvement.

G. Comparison to Other Methods

In this section, the performance of the proposed sparse one-class kernel fusion regression approach (NNZ=5) is compared against other methods on four datasets. Note that the proposed approach does not utilise any PA samples for training (i.e. operates in a zero-shot attack scenario). Nevertheless, we provide comparisons between the proposed method and both one-class (not utilising PA training data) as well as multi-class approaches (utilising PA training data).

TABLE III

EFFECT OF TEMPORAL AGGREGATION IN TERMS OF ACER (%). R.M.:REPLAY-MOBILE; R.A.: REPLAY-ATTACK; M.M.: MSU-MFSD; O.N.: OULU-NPU; ALL: AVERAGE OVER ALL DATASETS

	R.M.	R.A.	M.M.	O.N.	All
Frame-Level	15.92	1.91	4.01	14.14	8.99
Video-Level (raw)	15.68	0.0	0.0	6.67	5.58
Video-Level (prob.)	13.64	0.0	0.0	6.25	4.97

In addition to other existing face PAD techniques in the literature, in order to provide further baseline performances and enable a more critical assessment of the proposed technique, three other kernel-based one-class classifiers are included for comparison. These are Support Vector Data Description (SVDD) [74], Kernel Principal Component Analysis (KPCA) [75] and the Gaussian Process (GP) [76]. For a fair comparison, the SVDD, KPCA and GP benefit from a similar kernel fusion strategy as that of the sparse kernel regression approach.

1) OULU-NPU: The results of a comparison between the proposed approach and some other methods on the fourth protocol of the OULU-NPU dataset are reported in Table IV. From the table, it may be observed that the average ACER of the proposed kernel Fisher null-space approach on the OULU-NPU dataset (6.25 ± 6.85) is better than other unseen methods including SVDD, KPCA and GP. Compared to the multi-class methods, the proposed approach outperforms the majority of the existing methods in Table IV (except for [77]) that use PA data for training. As an instance, the proposed approach performs better than the FAS-BAS approach [20] with an ACER of 9.5 ± 6.0. The GRADIANT method [78] with an ACER of 10.0 ± 5.0 represents the best performing method on the fourth protocol of the OULU-NPU dataset in the "Competition on Generalized Software-based Face Presentation Attack Detection in Mobile Scenarios" [78].

2) Replay-Attack: The results of a comparison between the proposed approach and other methods in an unseen attack detection scenario on the Replay-Attack dataset are presented in Table V. As it may be observed from the table, the proposed approach achieves perfect detection performance on this dataset (AUC = 100%) while the best unseen face PAD method from the literature obtains a detection performance of 99.8% in terms of AUC (AUC is chosen as a the performance metric to enable a comparison with other methods). The proposed method also performs better than the SVDD approach while yielding a similar performance as those of KPCA and GP which both benefit from a similar multiple kernel fusion strategy.

Table VI presents a comparison between the state of the art multi-class methods and the proposed approach on the Replay-Attack dataset in terms of HTER. The proposed approach achieves a zero HTER, outperforming the majority of

TABLE IV

COMPARISON OF THE PERFORMANCE OF THE PROPOSED APPROACH TO OTHER METHODS (INCLUDING MULTI-CLASS METHODS) ON PROTOCOL 4 OF THE OULU-NPU DATASET. (SVDD, KPCA AND GP BENEFIT FROM THE SAME MULTIPLE KERNEL FUSION STRATEGY AS THAT OF THIS WORK.)

Method	APCER(%)	BPCER(%)	ACER (%)
Massy HNU [78]	35.8±35.3	8.3±4.1	22.1±17.6
GRADIANT[78]	5.0±4.5	15.0±7.1	10.0±5.0
FAS-BAS[20]	9.3±5.6	10.4±6.0	9.5±6.0
LBP-SVM[62]	41.67±27.03	55.0±21.21	48.33±6.07
IQM-SVM[62]	34.17±25.89	39.17±23.35	36.67±12.13
DeepPixBiS[62]	36.67±29.67	13.33±16.75	25.0±12.67
the work in [77]	0.9±1.8	4.2±5.3	2.6±2.8
SVDD	25.0±17.32	8.33±6.83	16.67±10.68
KPCA	13.33±14.72	11.67±11.25	12.5±12.94
GP	15.83±16.25	2.5±4.18	9.17±8.76
This work	11.67±13.66	0.83±2.04	6.25±6.85

TABLE V

COMPARISON OF THE PERFORMANCE OF THE PROPOSED APPROACH TO OTHER METHODS ON THE REPLAY-ATTACK DATASET IN AN UNSEEN ATTACK SCENARIO IN TERMS OF AUC (%).(SVDD, KPCA AND GP BENEFIT FROM THE SAME MULTIPLE KERNEL FUSION STRATEGY AS THAT OF THIS WORK.)

Method	AUC (%)
OCSVM+IMQ [7]	80.76
OCSVM+BSIF [7]	81.94
NN+LBP [9]	91.26
GMM+LBP [9]	90.06
OCSVM+LBP [9]	87.90
AE+LBP [9]	86.12
DTL [11]	99.80
One-Class MD [10]	99.75
SVDD	97.50
KPCA	100
GP	100
This work	100

TABLE VI

COMPARISON OF THE PERFORMANCE OF PROPOSED APPROACH TO THE STATE-OF-THE-ART MULTI-CLASS METHODS ON THE REPLAY-ATTACK DATASET IN TERMS OF HTER (%)

Method	HTER(%)
Boulkenafet et al. [80]	2.90
lsCNN [41]	2.50
lsCNN Traditionally Trained [41]	1.75
Chingovska et al. [81]	6.29
Boulkenafet et al. [36]	3.50
LBP + GS-LBP [29]	3.13
Patch-CNN [82]	1.25
Depth-CNN [82]	0.75
Patch&Depth-CNNs [82]	0.72
Image Quality [83]	0.03
Deep Learning [79]	0
The work in [49]	1.45
The work in [84]	3.95
This work	0

multi-class methods. The only method among others achieving a zero HTER is that of [79].

3) MSU-MFSD: Table VII presents a comparison between the proposed approach and other methods operating in an unseen PAD scenario in terms of AUC. The following observations may be made from the table. The purposed method performs better than other existing face PAD methods, achieving a perfect detection performance as compared to an AUC of 93% for the DTL method [11]. Compared with SVDD, KPCA and GP, the proposed solution performs better than SVDD, while KPCA and GP, benefiting from a similar multiple kernel fusion strategy, achieve a similar performance as that of the proposed approach.

TABLE VII

COMPARISON OF THE PERFORMANCE OF THE PROPOSED APPROACH TO OTHER METHODS ON THE MSU-MFSD DATASET IN AN UNSEEN ATTACK SCENARIO IN TERMS OF AUC (%). (SVDD, KPCA AND GP BENEFIT FROM THE SAME MULTIPLE KERNEL FUSION STRATEGY AS THAT OF THIS WORK.)

Method	AUC(%)
OCSVM+IMQ [7]	67.77
OCSVM+BSIF [7]	75.64
NN+LBP [9]	81.59
GMM+LBP [9]	81.34
OCSVM+LBP [9]	84.47
AE+LBP [9]	87.63
DTL [11]	93.00
SVDD	97.5
KPCA	100
GP	100
This work	100

TABLE VIII

COMPARISON OF THE PERFORMANCE OF PROPOSED APPROACH TO THE STATE-OF-THE-ART MULTI-CLASS METHODS ON THE MSU-MFSD DATASET IN TERMS OF EER (%)

Method	EER (%)
Texture analysis [80]	4.9
HSV+YCbCr fusion [86]	2.2
Multiscale Fusion [87]	6.9
IDA [73]	8.5
Colour LBP [36]	10.8
HRLF [88]	0.04
RD [85]	0.0
This work	0.0

The results of a comparison between this work and the state of the art multi-class methods in terms of EER are presented in Table VIII. As it may be observed from the table, the proposed approach (achieving a perfect detection performance) outperforms many multi-class methods. Among others, only the RD method [85] provides a similar performance.

4) Replay-Mobile: A similar comparison as those on other datasets is performed on the Replay-Mobile dataset. The results are reported in Table IX in an unseen PAD scenario and compared against others in terms of HTER (HTER is chosen for comparison as the majority of the existing unseen PAD results on this dataset have reported their performances in terms HTER). From the table, it may be observed that the proposed approach performs better than other methods in an unseen PAD scenario reported in the literature, achieving a HTER of 13.64% compared to the best performing unseen method of [10] with a HTER of 14.34%. The proposed one-class method also performs better than SVDD, KPCA and GP approaches for face PAD in an unseen attack scenario.

A comparison of the proposed approach to the state of the art multi-class methods is provided in Table X. As it may be observed from the table, the proposed method performs better

TABLE IX

COMPARISON OF THE PERFORMANCE OF THE PROPOSED APPROACH TO OTHER METHODS ON THE REPLAY-MOBILE DATASET IN AN UNSEEN ATTACK SCENARIO IN TERMS OF HTER (%).(SVDD, KPCA AND GP BENEFIT FROM THE SAME MULTIPLE KERNEL FUSION STRATEGY AS THAT OF THIS WORK.)

Method	**HTER** (%)
GoogleNet+SVDD [10]	14.34
ResNet50+SVDD [10]	21.76
VGG16+SVDD [10]	18.78
GoogleNet+MD [10]	13.70
ResNet50+MD [10]	21.81
VGG16+MD [10]	19.84
GoogleNet+GMM [10]	14.21
ResNet50+GMM [10]	21.53
VGG16+GMM [10]	18.05
SVDD	16.14
KPCA	17.05
GP	16.36
This work	11.88

TABLE X

COMPARISON OF THE PERFORMANCE OF PROPOSED APPROACH TO THE STATE-OF-THE-ART MULTI-CLASS METHODS ON THE REPLAY-MOBILE DATASET IN TERMS OF HTER (%)

Method	**HTER** (%)
two-class SVM + LBP [8]	17.2
two-class SVM + Motion [8]	10.4
two-class SVM + Gabor [8]	9.13
two-class SVM + IQM [8]	4.10
This work	11.88

than some other multi-class methods, yet it performs inferior compared to some multi-class techniques.

H. Summary of Performance Evaluations

Based on the evaluations conducted on four datasets, the proposed approach, when compared to other methods operating in an unseen attack scenario (not using PA samples for training) obtains the-state-of-the-art performance. Moreover, the performance of the proposed approach is also very competitive to the state-of-the-art multi-class methods. In this context, on two out of four datasets, the proposed approach achieves better or similar performance compared to methods that benefit from PA training samples for training.

The ROC curves corresponding to the proposed approach on different datasets are presented in Fig. 2. While on the Replay-Attack and MSU-MFSD datasets a perfect detection rate is obtained for all subjects, on the Replay-Mobile and Oulu-NPU datasets one observes that the performances of different subjects may vary to some extent. As an instance, for some subjects from the Oulu-NPU dataset a zero error rate is obtained while for the worst performing subject, the error rate could be around 30%. Similarly, on the Replay-Mobile dataset, while for some subjects perfect detection is achieved, for the worst performing subject, the error rate may be more than 28%.

In Fig. 3, the subjects with the worst detection rates from the Oulu-NPU and Replay-Mobile datasets are depicted. While the environmental imaging conditions do impact the detection, a common attribute among the subjects with the lowest detection

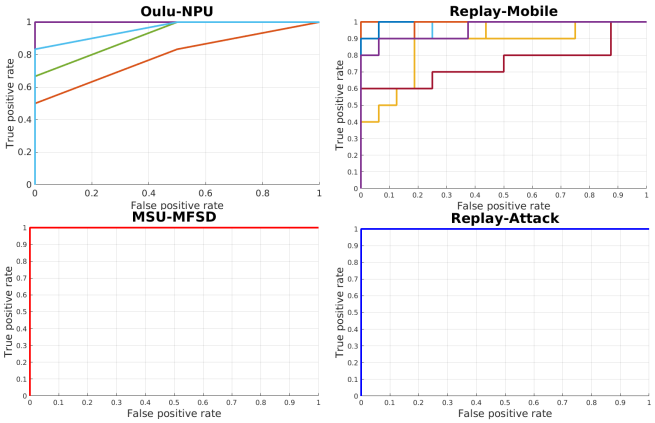


Fig. 2. ROC curves of the proposed approach on different datasets for different subjects (curves of clients with similar performances overlap).

Fig. 3. Subjects with the highest detection error rates from the Oulu-NPU (left, ACER = 33.3%) and Replay-Mobile (right, HTER = 28.1%) datasets.

rates may be that of eyeglasses with prominent frames. In this respect, the low detection rates associated with these subjects may be attributed to distinctive eyeglasses which could potentially alter the frequency information content of images (in addition to possible specular reflections) as a result of strong edges in the image in a way that the frequency content of presentation attacks may resemble more that of bona fide samples, and potentially be reflected in the features derived from (local) images. A deep investigation of this aspect of the method is flagged as a future direction investigation.

I. A Note on Inter-Dataset Evaluation

As discussed previously, the proposed face PAD technique based on the sparse one-class kernel fusion regression operates in a class-specific framework. For this purpose, subject-specific data is required to build client-specific classifiers. However, as there is not overlap between subjects from different datasets, the client-specific approach prevents an inter-dataset evaluation. While the cross-dataset evaluation is expected to be more challenging than the intra-dataset evaluation, it should be noted that the current study addresses a different, and possibly more challenging aspect of the face PAD problem, i.e. the unseen attack detection problem. In this respect, the difficulties associated with an inter-dataset evaluation attributed to, for instance, different imaging sensors, different illumination conditions, etc. are addressed in the experiments conducted on different datasets such as the OULU-NPU database which naturally incorporate such variations.

J. Remarks

- In a novelty detection task where no information regarding novel samples is accessible, it is challenging to

set the decision threshold. A common practice in this case is to set the decision threshold at a pre-specified confidence level such that a desired small proportion of positive samples is rejected. In the case of availability of a separate development set specific for each subject, one would be able to set the threshold such that a desired trade-off between FAR and FRR is maintained. In the anomaly approach followed in this work, the HTER corresponds to the confidence level point closest to the EER point. However, setting the decision threshold in an OCC approach to ensure a suitable practical performance remains an open problem, subject to future investigation.

- An important practical aspect of a face PAD approach is that of computational complexity. The recent trend in the field is directed towards utilising effective deep convolutional architectures. We have follow this approach but with two distinctions. First, the current study uses pre-trained networks. As such, it completely circumvents the difficulties associated with over-fitting due to small face PAD sample size or the computationally demanding training stage of the network. Second, instead of a single deep CNN, we have applied multiple networks not only to the whole face image but also to different local regions of the images. While understandably such a procedure increases the computational load as compared with a single network applied only to the whole face region, as demonstrated in the experiments, the kernel combination of multiple representations is effective in improving the detection performance. In this direction and in order to moderate the computational complexity at the operational phase, we developed a sparse variant of the proposed approach that effectively improves the computational complexity at the inference stage. The impact of such a sparse approach and its running times were studied in §V-E.
- In the current study, four different regional representations and three networks are used to construct multiple kernels. A further path of future investigation might be to examine different regional representations or to design a one-class Fisher null-space multiple kernel learning approach to enhance detection performance. In an ideal scenario, a face PAD system should be able to benefit from any previously seen presentation attacks. This may be done by refining the solution of a one-class classifier by possibly cutting off part of the feature subspace where some presentation attacks (anomalies) have been observed. In practice, however, the successful one-class anomaly-based approaches in the literature typically are unable to make use of any previously known anomalies. As an instance, the one-class kernel principal component (KPCA), the Gaussian Process (PG) and the Kernel Fisher Null-Space Regression approach do not provide a direct and mathematically sound mechanism to benefit from any possible negative training observations. As a future direction of investigation, we are planning to extend the methodology presented in this study so that it would be able to make use of any seen presentation attacks to refine the solution of a one-class classifier.

VI. Conclusion

The paper addressed the face presentation attack detection problem in the challenging conditions of an unseen attack detection setting. For this purpose, a one-class novelty detection approach, based on kernel regression was presented. Benefiting from generic deep CNN representations, additional mechanisms including a multiple kernel fusion approach, sparse regularisation of the regression solution, client-specific, and probabilistic modelling were introduced to improve the performance of the system. Experimental evaluation of the proposed approach on four publicly available datasets in an unseen face PAD setting illustrated that the proposed method outperforms other methods operating in an unseen scenario while competing closely with the state-of-the-art multi-class methods. In the case of availability of PA training samples, further improvements may be achieved by refining the solution of a one-class learner through different mechanisms including, for instance, a feature selection procedure, or multiple kernel learning using multi-class training samples, etc. which may be consiered as future directions for investigation.

References

[1] C.-Y. Low, A. B.-J. Teoh, and C.-J. Ng, "Multi-fold Gabor, PCA, and ICA filter convolution descriptor for face recognition," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 29, no. 1, pp. 115–129, Jan. 2019.

[2] A. Sepas-Moghaddam, M. A. Haque, P. L. Correia, K. Nasrollahi, T. B. Moeslund, and F. Pereira, "A double-deep spatio-angular learning framework for light field-based face recognition," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 30, no. 12, pp. 4496–4512, Dec. 2020.

[3] H. Cevikalp, H. S. Yavuz, and B. Triggs, "Face recognition based on videos by using convex hulls," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 30, no. 12, pp. 4481–4495, Dec. 2020.

[4] S. Yang, W. Deng, M. Wang, J. Du, and J. Hu, "Orthogonality loss: Learning discriminative representations for face recognition," *IEEE Trans. Circuits Syst. Video Technol.*, early access, Sep. 2, 2020, doi: 10.1109/TCSVT.2020.3021128.

[5] S. Bhattacharjee, A. Mohammadi, A. Anjos, and S. Marcel, *Recent Advances in Face Presentation Attack Detection*. Cham, Switzerland: Springer, 2019, pp. 207–228.

[6] S. Marcel, M. S. Nixon, J. Fiérrez, and N. W. D. Evans, Eds., *Handbook Biometric Anti-Spoofing: Presentation Attack Detection* (Advances in Computer Vision and Pattern Recognition), 2nd ed. Cham, Switzerland: Springer, 2019.

[7] S. R. Arashloo, J. Kittler, and W. Christmas, "An anomaly detection approach to face spoofing detection: A new formulation and evaluation protocol," *IEEE Access*, vol. 5, pp. 13868–13882, 2017.

[8] O. Nikisins, A. Mohammadi, A. Anjos, and S. Marcel, "On effectiveness of anomaly detection approaches against unseen presentation attacks in face anti-spoofing," in *Proc. Int. Conf. Biometrics (ICB)*, Feb. 2018, pp. 75–81.

[9] F. Xiong and W. Abdalmageed, "Unknown presentation attack detection with face RGB images," in *Proc. IEEE 9th Int. Conf. Biometrics Theory, Appl. Syst. (BTAS)*, Oct. 2018, pp. 1–9.

[10] S. Fatemifar, S. R. Arashloo, M. Awais, and J. Kittler, "Spoofing attack detection by anomaly detection," in *Proc. IEEE Int. Conf. Acoust., Speech Signal Process. (ICASSP)*, May 2019, pp. 8464–8468.

[11] Y. Liu, J. Stehouwer, A. Jourabloo, and X. Liu, "Deep tree learning for zero-shot face anti-spoofing," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Long Beach, CA, USA, 2019, pp. 4675–4684.

[12] D. Perez-Cabo, D. Jimenez-Cabello, A. Costa-Pazo, and R. J. Lopez-Sastre, "Deep anomaly detection for generalized face anti-spoofing," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. Workshops (CVPRW)*, Jun. 2019, pp. 1–10.

[13] A. Rattani and A. Ross, "Automatic adaptation of fingerprint liveness detector to new spoof materials," in *Proc. IEEE Int. Joint Conf. Biometrics*, Sep. 2014, pp. 1–8.

[14] A. Rattani, W. J. Scheirer, and A. Ross, "Open set fingerprint spoof detection across novel fabrication materials," *IEEE Trans. Inf. Forensics Security*, vol. 10, no. 11, pp. 2447–2460, Nov. 2015.

[15] 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.

[16] P. Tuveri, M. Zurutuza, and G. L. Marcialis, "Incremental support vector machine for self-updating fingerprint presentation attack detection systems," in *Image Analysis and Processing*, S. Battiato, G. Gallo, R. Schettini, and F. Stanco, Eds. Cham, Switzerland: Springer, 2017, pp. 739–749.

[17] P. M. Ferreira, A. F. Sequeira, D. Pernes, A. Rebelo, and J. S. Cardoso, "Adversarial learning for a robust iris presentation attack detection method against unseen attack presentations," in *Proc. Int. Conf. Biometrics Special Interest Group (BIOSIG)*, Sep. 2019, pp. 1–7.

[18] M. Sahidullah *et al.*, "Introduction to voice presentation attack detection and recent advances," in *Handbook of Biometric Anti-Spoofing*. Cham, Switzerland: Springer, 2019, pp. 321–361.

[19] Y. Zhang, X. Tian, Y. Li, X. Wang, and D. Tao, "Principal component adversarial example," *IEEE Trans. Image Process.*, vol. 29, pp. 4804–4815, 2020.

[20] Y. Liu, A. Jourabloo, and X. Liu, "Learning deep models for face anti-spoofing: Binary or auxiliary supervision," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit.*, Jun. 2018, pp. 389–398.

[21] Y. Liu, J. Stehouwer, A. Jourabloo, and X. Liu, "Deep tree learning for zero-shot face anti-spoofing," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2019, pp. 4680–4689.

[22] S. S. Khan and M. G. Madden, "One-class classification: Taxonomy of study and review of techniques," *Knowl. Eng. Rev.*, vol. 29, no. 3, pp. 345–374, Jun. 2014.

[23] J. Villalba, A. Miguel, A. Ortega, and E. Lleida, "Spoofing detection with DNN and one-class SVM for the ASVspoof 2015 challenge," in *Proc. 16th Annu. Conf. Int. Speech Commun. Assoc.*, Sep. 2015, pp. 2067–2071.

[24] A. F. Sequeira, S. Thavalengal, J. Ferryman, P. Corcoran, and J. S. Cardoso, "A realistic evaluation of iris presentation attack detection," in *Proc. 39th Int. Conf. Telecommun. Signal Process. (TSP)*, Jun. 2016, pp. 660–664.

[25] J. Kittler, M. Hatef, R. P. W. Duin, and J. Matas, "On combining classifiers," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 20, no. 3, pp. 226–239, Mar. 1998.

[26] T. Edmunds, "Protection of 2D face identification systems against spoofing attacks," M.S. thesis, Univ. Grenoble Alpes, Grenoble, France, Jan. 2017.

[27] R. Tolosana, M. Gomez-Barrero, C. Busch, and J. Ortega-Garcia, "Biometric presentation attack detection: Beyond the visible spectrum," *IEEE Trans. Inf. Forensics Security*, vol. 15, pp. 1261–1275, 2020.

[28] Z. Boulkenafet, J. Komulainen, and A. Hadid, "On the generalization of color texture-based face anti-spoofing," *Image Vis. Comput.*, vol. 77, pp. 1–9, Sep. 2018.

[29] F. Peng, L. Qin, and M. Long, "Face presentation attack detection using guided scale texture," *Multimedia Tools Appl.*, vol. 77, no. 7, pp. 8883–8909, Apr. 2018.

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

[31] L. Feng *et al.*, "Integration of image quality and motion cues for face anti-spoofing: A neural network approach," *J. Vis. Commun. Image Represent.*, vol. 38, pp. 451–460, Jul. 2016.

[32] A. Pinto, H. Pedrini, W. R. Schwartz, and A. Rocha, "Face spoofing detection through visual codebooks of spectral temporal cubes," *IEEE Trans. Image Process.*, vol. 24, no. 12, pp. 4726–4740, Dec. 2015.

[33] G. Kim, S. Eum, J. K. Suhr, D. I. Kim, K. R. Park, and J. Kim, "Face liveness detection based on texture and frequency analyses," in *Proc. 5th IAPR Int. Conf. Biometrics (ICB)*, Mar. 2012, pp. 67–72.

[34] A. Pinto, W. R. Schwartz, H. Pedrini, and A. De Rezende Rocha, "Using visual rhythms for detecting video-based facial spoof attacks," *IEEE Trans. Inf. Forensics Security*, vol. 10, no. 5, pp. 1025–1038, May 2015.

[35] J. Galbally and S. Marcel, "Face anti-spoofing based on general image quality assessment," in *Proc. 22nd Int. Conf. Pattern Recognit.*, Aug. 2014, pp. 1173–1178.

[36] Z. Boulkenafet, J. Komulainen, and A. Hadid, "Face anti-spoofing based on color texture analysis," in *Proc. IEEE Int. Conf. Image Process. (ICIP)*, Sep. 2015, pp. 2636–2640.

[37] T. Wang, J. Yang, Z. Lei, S. Liao, and S. Z. Li, "Face liveness detection using 3D structure recovered from a single camera," in *Proc. Int. Conf. Biometrics (ICB)*, Jun. 2013, pp. 1–6.

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

[39] H. P. Nguyen, A. Delahaies, F. Retraint, and F. Morain-Nicolier, "Face presentation attack detection based on a statistical model of image noise," *IEEE Access*, vol. 7, pp. 175429–175442, 2019.

[40] X. Zhu, S. Li, X. Zhang, H. Li, and A. C. Kot, "Detection of spoofing medium contours for face anti-spoofing," *IEEE Trans. Circuits Syst. Video Technol.*, early access, Oct. 28, 2019, doi: 10.1109/TCSVT.2019.2949868.

[41] G. B. de Souza, J. P. Papa and A. N. Marana, "On the learning of deep local features for robust face spoofing detection," in *Proc. 31st SIBGRAPI Conf. Graph., Patterns Images (SIBGRAPI)*, Parana, Argentina, 2018, pp. 258–265.

[42] J. Yang, Z. Lei, and S. Z. Li, "Learn convolutional neural network for face anti-spoofing," 2014, *arXiv:1408.5601*. [Online]. Available: https://arxiv.org/abs/1408.5601

[43] A. George, Z. Mostaani, D. Geissenbuhler, O. Nikisins, A. Anjos, and S. Marcel, "Biometric face presentation attack detection with multi-channel convolutional neural network," *IEEE Trans. Inf. Forensics Security*, vol. 15, pp. 42–55, 2020.

[44] G. Heusch and S. Marcel, "Remote blood pulse analysis for face presentation attack detection," in *Handbook of Biometric Anti-Spoofing*. Cham, Switzerland: Springer, 2019, pp. 267–289.

[45] L. Li, X. Feng, Z. Xia, X. Jiang, and A. Hadid, "Face spoofing detection with local binary pattern network," *J. Vis. Commun. Image Represent.*, vol. 54, pp. 182–192, Jul. 2018.

[46] B. Hamdan and K. Mokhtar, "The detection of spoofing by 3D mask in a 2D identity recognition system," *Egyptian Informat. J.*, vol. 19, no. 2, pp. 75–82, Jul. 2018.

[47] T. Edmunds and A. Caplier, "Motion-based countermeasure against photo and video spoofing attacks in face recognition," *J. Vis. Commun. Image Represent.*, vol. 50, pp. 314–332, Jan. 2018.

[48] M. Killioğlu, M. Taşkiran, and N. Kahraman, "Anti-spoofing in face recognition with liveness detection using pupil tracking," in *Proc. IEEE 15th Int. Symp. Appl. Mach. Intell. Inform. (SAMI)*, Jan. 2017, pp. 87–92.

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

[50] E. E. A. Abusham and H. K. Bashir, "Face recognition using local graph structure (LGS)," in *Human-Computer Interaction. Interaction Techniques and Environments*, J. A. Jacko, Ed. Berlin, Germany: Springer, 2011, pp. 169–175.

[51] D. Caetano Garcia and R. L. de Queiroz, "Face-spoofing 2D-detection based on Moiré-pattern analysis," *IEEE Trans. Inf. Forensics Security*, vol. 10, no. 4, pp. 778–786, Apr. 2015.

[52] R. Ramachandra and C. Busch, "Presentation attack detection methods for face recognition systems: A comprehensive survey," *ACM Comput. Surv.*, vol. 50, no. 1, pp. 8:1–8:37, Mar. 2017.

[53] B. Schölkopf, J. C. Platt, J. 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, Jul. 2001.

[54] S. R. Arashloo and J. Kittler, "Robust one-class kernel spectral regression," *IEEE Trans. Neural Netw. Learn. Syst.*, early access, Apr. 2, 2020, doi: 10.1109/TNNLS.2020.2979823.

[55] E. Alpaydin, *Introduction to Machine Learning* (Adaptive Computation and Machine Learning), 3rd ed. Cambridge, MA, USA: MIT Press, 2014.

[56] J. Liu, Z. Lian, Y. Wang, and J. Xiao, "Incremental kernel null space discriminant analysis for novelty detection," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jul. 2017, pp. 4123–4131.

[57] P. Bodesheim, A. Freytag, E. Rodner, M. Kemmler, and J. Denzler, "Kernel null space methods for novelty detection," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Jun. 2013, pp. 3374–3381.

[58] 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.

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

[60] 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.

[61] F. Yan, J. Kittler, K. Mikolajczyk, and A. Tahir, "Non-sparse multiple kernel Fisher discriminant analysis," *J. Mach. Learn. Res.*, vol. 13, no. 1, pp. 607–642, Mar. 2012.

[62] A. George and S. Marcel, "Deep pixel-wise binary supervision for face presentation attack detection," in *Proc. Int. Conf. Biometrics (ICB)*, Jun. 2019, pp. 1–8.

[63] C. Szegedy *et al.*, "Going deeper with convolutions," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2015, pp. 1–9.

[64] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2016, pp. 770–778.

[65] K. Simonyan and A. Zisserman, "Very deep convolutional networks for large-scale image recognition," in *Proc. 3rd Int. Conf. Learn. Represent. (ICLR)*, San Diego, CA, USA, May 2015.

[66] B. Efron, T. Hastie, I. Johnstone, and R. Tibshirani, "Least angle regression," *Ann. Statist.*, vol. 32, no. 2, pp. 407–451, 2004.

[67] S. Fatemifar, S. R. Arashloo, M. Awais, and J. Kittler, "Client-specific anomaly detection for face presentation attack detection," *Pattern Recognit.*, Oct. 2020, Art. no. 107696.

[68] V. Štruc and N. Pavešić, *Photometric Normalization Techniques for Illumination Invariance*. Hershey, PA, USA: IGI-Global, 2011, pp. 279–300.

[69] B. Amos, B. Ludwiczuk, and M. Satyanarayanan, "OpenFace: A general-purpose face recognition library with mobile applications," CMU School Comput. Sci., Pittsburgh, PA, USA, Tech. Rep. CMU-CS-16-118, 2016.

[70] A. Costa-Pazo, S. Bhattacharjee, E. Vazquez-Fernandez, and S. Marcel, "The replay-mobile face presentation-attack database," in *Proc. Int. Conf. Biometrics Special Interest Group (BIOSIG)*, Sep. 2016, pp. 1–7.

[71] 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.

[72] Z. Boulkenafet, J. Komulainen, L. Li, X. Feng, and A. Hadid, "OULU-NPU: A mobile face presentation attack database with real-world variations," in *Proc. 12th IEEE Int. Conf. Automat. Face Gesture Recognit. (FG)*, May 2017, pp. 612–618.

[73] 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.

[74] D. M. J. Tax and R. P. W. Duin, "Support vector data description," *Mach. Learn.*, vol. 54, no. 1, pp. 45–66, Jan. 2004.

[75] H. Hoffmann, "Kernel PCA for novelty detection," *Pattern Recognit.*, vol. 40, no. 3, pp. 863–874, Mar. 2007.

[76] M. Kemmler, E. Rodner, E.-S. Wacker, and J. Denzler, "One-class classification with Gaussian processes," *Pattern Recognit.*, vol. 46, no. 12, pp. 3507–3518, Dec. 2013.

[77] X. Yang *et al.*, "Face anti-spoofing: Model matters, so does data," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2019, pp. 3502–3511.

[78] Z. Boulkenafet *et al.*, "A competition on generalized software-based face presentation attack detection in mobile scenarios," in *Proc. IEEE Int. Joint Conf. Biometrics (IJCB)*, Oct. 2017, pp. 688–696.

[79] I. Manjani, S. Tariyal, M. Vatsa, R. Singh, and A. Majumdar, "Detecting silicone mask-based presentation attack via deep dictionary learning," *IEEE Trans. Inf. Forensics Security*, vol. 12, no. 7, pp. 1713–1723, Jul. 2017.

[80] 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.

[81] I. Chingovska, "Trustworthy biometric verification under spoofing attacks: Application to the face mode," Ph.D. dissertation, École Polytechnique Fédérale de Lausanne, Lausanne, Switzerland, 2016, no. 6879.

[82] Y. Atoum, Y. Liu, A. Jourabloo, and X. Liu, "Face anti-spoofing using patch and depth-based CNNs," in *Proc. IEEE Int. Joint Conf. Biometrics (IJCB)*, Oct. 2017, pp. 319–328.

[83] E. Fourati, W. Elloumi, and A. Chetouani, "Anti-spoofing in face recognition-based biometric authentication using image quality assessment," *Multimedia Tools Appl.*, vol. 79, nos. 1–2, pp. 865–889, Oct. 2019.

[84] I. Chingovska and A. Rabello 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.

[85] T. Edmunds and A. Caplier, "Fake face detection based on radiometric distortions," in *Proc. 6th Int. Conf. Image Process. Theory, Tools Appl. (IPTA)*, Dec. 2016, pp. 1–6.

[86] Z. Boulkenafet, J. Komulainen, and A. Hadid, "Face antispoofing using speeded-up robust features and Fisher vector encoding," *IEEE Signal Process. Lett.*, vol. 24, no. 2, pp. 141–145, Feb. 2017.

[87] Z. Boulkenafet, J. Komulainen, X. Feng, and A. Hadid, "Scale space texture analysis for face anti-spoofing," in *Proc. Int. Conf. Biometrics (ICB)*, Jun. 2016, pp. 1–6.

[88] U. Muhammad and A. Hadid, "Face anti-spoofing using hybrid residual learning framework," in *Proc. Int. Conf. Biometrics (ICB)*, Jun. 2019, pp. 1–7.

Shervin Rahimzadeh Arashloo received the Ph.D. degree from the Center for Vision, Speech and Signal Processing (CVSSP), University of Surrey, Guildford, U.K., in 2010. He is currently an Assistant Professor with the Department of Computer Engineering, Bilkent University, Ankara, Turkey, and a Visiting Research Fellow with the CVSSP, University of Surrey. His research interests include pattern recognition, machine learning, and signal processing.