Unsupervised Domain Adaptation for Face Anti-Spoofing

Haoliang Li, *Student Member, IEEE*, Wen Li, *Member, IEEE*, Hong Cao, *Senior Member, IEEE*, Shiqi Wang, *Member, IEEE*, Feiyue Huang, and Alex C. Kot, *Fellow, IEEE*

***Abstract*— Face anti-spoofing (a.k.a. presentation attack detection) has recently emerged as an active topic with great significance for both academia and industry due to the rapidly increasing demand in user authentication on mobile phones, PCs, tablets, and so on. Recently, numerous face spoofing detection schemes have been proposed based on the assumption that training and testing samples are in the same domain in terms of the feature space and marginal probability distribution. However, due to unlimited variations of the dominant conditions (illumination, facial appearance, camera quality, and so on) in face acquisition, such single domain methods lack generalization capability, which further prevents them from being applied in practical applications. In light of this, we introduce an unsupervised domain adaptation face anti-spoofing scheme to address the real-world scenario that learns the classifier for the target domain based on training samples in a different source domain. In particular, an embedding function is first imposed based on source and target domain data, which maps the data to a new space where the distribution similarity can be measured. Subsequently, the Maximum Mean Discrepancy between the latent features in source and target domains is minimized such that a more generalized classifier can be learned. State-of-the-art representations including both hand-crafted and deep neural network learned features are further adopted into the framework to quest the capability of them in domain adaptation. Moreover, we introduce a new database for face spoofing detection, which contains more than 4000 face samples with a large variety of spoofing types, capture devices, illuminations, and so on. Extensive experiments on existing benchmark databases and the new database verify that the proposed approach can gain significantly better generalization capability in cross-domain scenarios by providing consistently better anti-spoofing performance.**

***Index Terms*— Face anti-spoofing, domain adaptation, maximum mean discrepancy.**

Manuscript received October 31, 2016; revised March 29, 2017, August 4, 2017, and November 12, 2017; accepted January 16, 2018. Date of publication February 2, 2018; date of current version March 27, 2018. The associate editor coordinating the review of this manuscript and approving it for publication was Prof. Yunhong Wang. *(Corresponding author: Haoliang Li.)*

H. Li and A. C. Kot are with the School of Electrical and Electronic Engineering, Nanyang Technological University, Singapore 639798 (e-mail: hli016@e.ntu.edu.sg; eackot@ntu.edu.sg).

W. Li is with the Computer Vision Laboratory, ETH Zürich, 8092 Zürich, Switzerland (e-mail: liwen@vision.ee.ethz.ch).

H. Cao is with Ernst & Young Advisory Pte Ltd., Singapore (e-mail: hong.cao@sg.ey.com).

S. Wang is with the Department of Computer Science, City University of Hong Kong, Hong Kong (e-mail: shiqwang@cityu.edu.hk).

F. Huang is with the Tencent Youtu Laboratory, Shanghai 200233, China (e-mail: garyhuang@tencent.com).

Color versions of one or more of the figures in this paper are available online at http://ieeexplore.ieee.org.

Digital Object Identifier 10.1109/TIFS.2018.2801312

I. Introduction

FACE verification, which aims to authenticate a claimed identity based on the captured face image/video and make a decision to either accept or reject the request based on the matching result, has received more and more attention recently. Compared with the traditional authentication system which makes use of user's one-stroke draw pattern and password, face verification has the unique advantage to verify a person since the traditional passwords can be easily stolen and used by an attacker. Moreover, since the face recognition is non-intrusive and face images can be feasibly obtained with digital devices, face verification has been widely applied in various areas such as information security and access control.

Although face recognition is a challenging problem, many algorithms have recently been proposed with great success to lead it to be a more mature field of research. However, the face verification system is easily bypassed by a fake face image/video [1], [2]. In the trend of rapid proliferation of internet images, face spoofing becomes easier by means of LinkedIn, Facebook, and Webcam chat software (e.g. QQ, Skype). Therefore, the capability of filtering off the fake face is urgently required to allay the security concerns.

Generally speaking, face spoofing mainly consists of photo, masking, video and 3D attacks. For photo attack, a face image is firstly reproduced on a high-quality paper or displayed on the screen of a digital device, which is subsequently presented in front of a capturing camera for verification. Besides the printed paper and screen display, an advanced attack method is the masking attack with cut eyes and mouth, which was introduced in [3]. Video attack refers to displaying a face video recorded by a digital display device, e.g., tablet and notebook, for verification. Compared with standard photo attack, masking and video attacks are more sophisticated since they can introduce the motion and liveness information to improve the sense of reality. 3D attack is based on the face model with the popularity of 3D printing technology [4] and virtual reality [5]. However, 3D attack is much more expensive to launch compared with the traditional photo, masking, and video attacks.

To tackle the face anti-spoofing problem, numerous approaches have been proposed where generic classifiers based on extracted spoofing features were learned. While good results have been demonstrated on benchmark database, they were lack of generalization ability due to the assumption that training and testing face samples are captured under similar conditions. As observed in many works [6]–[13], the performance of face spoofing detection may drop significantly under the cross-database scenario. This greatly hinders the application of such methods in real application scenarios,

as given the testing samples it is always inconvenient to generate labeled training samples captured in similar conditions. In this paper, we focus on this problem and propose an unsupervised domain adaptation framework for face anti-spoofing to bridge the gap between the generic and domain adapted classifiers. To the best of our knowledge, there is no prior work imposing unsupervised domain adaptation to solve the face spoofing detection problem. In particular, given the labeled samples in source domain and unlabeled samples in target domain, it is practical for us to train a domain specific classifier based on inherent properties of these data.

There are three main contributions in our work:

- We cast the face anti-spoofing problem that learns a classifier from a different domain data into an unsupervised domain adaptation framework. The performance of the proposed scheme is evaluated in the cross-database scenario involving face data drawn from different conditions, and significantly better face spoofing detection performance has been observed based on our experimental results.
- State-of-the-art features are incorporated into the domain adaptation framework, and their performance, as well as the generalization ability, are analyzed. In particular, both hand-crafted features and deep neural network learned features are adopted, and their performances are demonstrated, analyzed and compared to quest the capability of these features in domain adaptation.
- We introduce a new face spoofing database. Compared with the existing databases, our new database covers more diverse camera models, lighting conditions and backgrounds with different attacking types. The face samples are captured by mobile phones with cameras ranging from high to low quality. In total, more than 4000 face videos have been collected.

The rest of this paper is organized as follows. In Section II, we provide a brief review regarding the related works on face spoofing detection as well as domain adaptation for biometrics. In Section III, our proposed domain adaptation framework is introduced. Then, we introduce the adopted features in Section IV. Experimental results are shown in Section V and Section VI concludes this paper.

II. Related Works

A. Face Spoofing Detection

In the past few years, numerous face spoofing detection techniques have been proposed. These techniques mainly focus on exploring efficient and effective features for face anti-spoofing which can be further divided into motion, texture, distortion and deep learning based approaches. Each category in its own way has made important contributions to face anti-spoofing. Generally speaking, motion based methods refer to extracting motion features such as optical flow for liveness detection. Texture based methods focus on adopting texture descriptors (e.g. Local Binary Pattern) as discriminative features. Distortion methods are based on distortion sensitive features. Deep learning based methods aim at learning the feature representation with the deep neural network for anti-spoofing.

1) Motion Based Methods: Motion based methods aim at extracting liveness information that can distinguish the genuine face from the fake one through eye blinking, lips movement, and head rotation. In [14], the planar object movements were detected as cues for translation, in-plane rotation, panning and out-of-plane rotation. In [15], the subtle movements of different facial parts were extracted as important features under the assumption that the genuine and fake faces can be distinguished by the movement cues. Furthermore, in [16]–[18] the background motion information was also utilized to detect face spoofing. However, though motion-based methods have achieved satisfactory performance for face spoofing detection, user cooperation is required for supplying the liveliness information for authentication such as user blinks eyes based on system's instruction. Moreover, extracting optical flow is to some extent time consuming, which limits their application for practical use on resource-constrained mobile platforms.

2) Texture Based Methods: To the best of our knowledge, the first work based on texture information for face anti-spoofing was proposed in [19], where two dimensional Fourier spectrum analysis was conducted. Tan *et al.* [20] proposed a difference-of-Gaussian (DoG) method to extract frequency disturbance based on the idea that a face image should be more severely distorted when it passes through the camera twice. Subsequently, the delicate multi-scale local binary patterns (LBP) [21] were employed in [22], which were encoded into an enhanced histogram for face anti-spoofing. By assuming that a fake face image may have different micro-textures compared with the genuine face, the authors adopted the multi-scale LBP as discriminative features to describe the micro-textures as well as their spatial information. The multi-scale LBP feature was further extended to Component Dependent based method to extract more discriminative information [23]. Other texture based methods by adopting Scale-invariant feature transform (SIFT), Speed-up Robust feature (SURF) and local phase quantization (LPQ) features were also discussed in [24] and [25]. Moreover, researchers extended the LBP into the three orthogonal planes (LBP-TOP [26], LPQ-TOP, BSIF-LBP [27]) to extract texture information in spatial and temporal domain based on the face videos. The recent works [8], [9], [28] have largely improved the spoofing detection performance by exploiting the joint color and texture information based on LBP and LPQ.

3) Distortion Based Methods: Methods in this category for face anti-spoofing are based on the inspirations that the fake images usually have lower quality than the genuine ones [29]. In [29], twenty-five image quality assessment metrics were adopted as features for learning the classifier. The employed metrics cover pixel-based and structure-based measures. A 3×3 low-pass Gaussian filter was applied to generate reference images for full-reference quality assessment.

Moreover, a distortion based face spoofing detection method was proposed in [7] by adopting four different features (specular reflection, blurriness, chromatic moment and color diversity). The specular reflection component was firstly extracted by [30], and then the statistical features based on specular component percentage, mean value of specular pixels and the variance were computed. For the blurriness feature, the no-reference blur score based on [31] and [32] were utilized. Regarding the chromatic moment and color diversity, the statistics features based on HSV and color quantization were adopted in [33].

4) Deep Learning Based Methods: Recently, convolutional neural network (CNN) was adopted for biometric spoofing detection. Nogueira *et al.* [34] showed that the pretrained network based on ImageNet [35] can be transferred to fingerprint spoofing detection scenario without fine-tuning process.

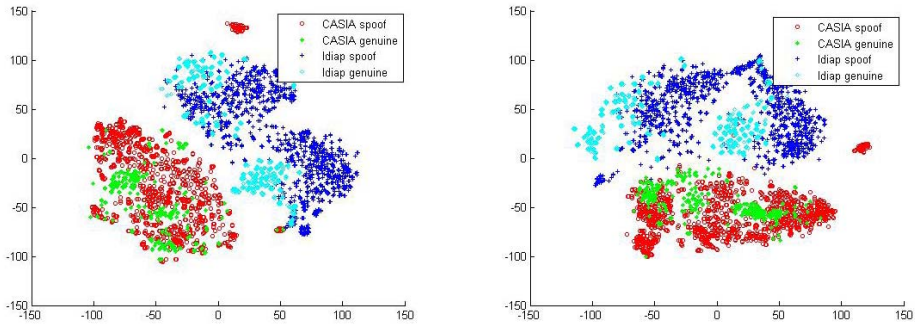

Fig. 1. 2D visualization of CASIA database and Idiap REPLAY-ATTACK database with different features. The figure is better to be viewed in color format. (a) Sample visualization with CoALBP feature. (b) Sample visualization with LPQ feature.

Menotti *et al.* [36] proposed a grid-search method to find a suitable model for biometric spoofing detection. Yang *et al.* [37] proposed to learn the CNN model based on the architecture of Krizhevsky *et al.* [38], which was proved to be effective for face spoofing detection.

5) Other Methods: Besides the methods mentioned above, some other techniques, including 3D depth [39], infrared (IR) image [40], voice [41], light-field camera image [42] and vein flow detection [43] were also proposed for face spoofing detection. However, these methods either need human interaction or extra equipment setup. For example, extra sensors are required to detect vein information in [43] and the speech analyzer is needed in [41].

While both hand-crafted feature based and deep learning based methods can achieve good performance based on intra-database scenario, large performance drop can still be observed in cross-database face spoofing detection scenario [6]–[13]. In [7], the authors proposed to employ distortion based feature which can be less influenced by the diverse facial appearance. In [9], the authors observed that transforming a face image to a new color space can improve the face spoofing detection performance. In our work, we focus on leveraging the advantage of domain adaptation to address the cross-database face spoofing detection problem.

B. Domain Adaptation

One common assumption in computer vision and machine learning is that the training data and testing data are sampled from the same distribution [44]. However, many practical scenarios (e.g. face verification and spoofing detection) involve data coming from different distributions (facial appearance and pose, illumination conditions, camera devices, etc.). Therefore, we may suffer from the overfitting problem with a significant performance drop when testing pre-trained model with the slightly different unseen data. Domain adaptation is associated with transfer learning which aims at solving the learning problem in target domain under a certain distribution with training data in source domain under another distribution. It has been extensively studied in recent computer vision tasks [45], [46]. For face verification, Cao *et al.* [47] proposed a semi-supervised transfer learning approach with joint Bayesian prior to deal with the challenge of different facial appearance distributions. For face spoofing detection, Yang *et al.* [13] proposed a subject based transformation method to synthesize fake face features based on the assumption that the relationship between genuine and fake samples belonging to an individual subject can be formulated as a linear transformation. However, in practice, the dominant factors in face capturing, including the camera models and illumination variations, can be even more diverse. This inspires us to introduce the unsupervised domain adaptation framework for face anti-spoofing.

III. Cross-Domain Face Spoofing Detection

In this section, the cross-domain face anti-spoofing technique is introduced to deal with the scenario that face samples for verification may not be taken by similar camera models or under similar illumination conditions to the training face samples. In general, the performance of the cross-domain face spoofing detection can directly reflect the generalization capability of the anti-spoofing algorithm. In particular, we focus on solving this problem from the perspective of machine learning. Inspired by the success of unsupervised domain adaptation in many applications [45], [46], we employ this technique to improve the generalization ability of face spoofing detection. The earlier work addressing the domain adaptation problem on face spoofing detection was reported in [13], where features of fake faces were synthesized by domain adaptation under the assumption that samples from different subjects can be formulated with a linear transformation. However, this method was conducted in a supervised manner by assuming the fixed linear transformations, which may not always hold in practice as face samples for verification can be captured by different camera models.

In this section, we assume that the source domain data (training samples) and target domain data (testing samples) preserve a certain statistical distribution. Fig. 1 visualizes the features[1] of face samples from different domains. We can observe that the features extracted from one particular domain tend to form a compact distribution, implying that they are taken under similar conditions. As such, unsupervised domain adaptation can be imposed to improve the generalization ability of face spoofing detection by minimizing the Maximum Mean Discrepancy (MMD) with which we adapt the pre-trained classifier from the source domain to target domain [49]–[51]. It is also worth mentioning that the proposed strategy is applicable in real-world scenarios, as it is feasible to collect unlabeled samples for face verification and anti-spoofing purpose via a variety of web services with photo information.

[1] We employ t-SNE [48] for feature visualization.

A. Maximum Mean Discrepancy Minimization

Maximum Mean Discrepancy (MMD) is defined as a distance metric for comparing two probability distributions (a.k.a. two-sample test) [49]. Typically, two distributions are identical if and only if the MMD distance equals to zero.

Given the facial feature samples $\mathbf{X}_s = [\mathbf{x}_{s1}; \mathbf{x}_{s2}; \ldots; \mathbf{x}_{sn_1}]$ (each row denotes a feature sample) which are used for training as source domain data and unlabeled face samples $\mathbf{X}_t = [\mathbf{x}_{t1}; \mathbf{x}_{t2}; \ldots; \mathbf{x}_{tn_2}]$ for verification as target domain data, MMD function which measures the distribution between two probability distributions can be defined as follows,

$$D(\mathbf{X}_s, \mathbf{X}_t) = \| \frac{1}{n_1} \sum_{i=1}^{n_1} \phi(\mathbf{x}_{si}) - \frac{1}{n_2} \sum_{i=1}^{n_2} \phi(\mathbf{x}_{ti}) \|^2 \tag{1}$$

where n_1 is the number of samples in source domain (for training), n_2 is the number of samples in target domain (for testing) and ϕ refers to the embedding function which maps the data from feature space to the Reproducing Kernel Hilbert Space (RKHS) where the distance between two probability distributions can be properly measured [49].

The MMD function can also be reformulated in matrix form by considering a kernel representation of source domain data and target domain data, which is given as

$$\mathbf{K} = \begin{pmatrix} \mathbf{K}_{ss} & \mathbf{K}_{st} \\ \mathbf{K}_{ts} & \mathbf{K}_{tt} \end{pmatrix} \in \mathbb{R}^{(n1+n2)\times(n1+n2)} \tag{2}$$

where $\mathbf{K}_{ss} = [\phi(\mathbf{x}_{si})\phi(\mathbf{x}_{sj})^T], \mathbf{K}_{tt} = [\phi(\mathbf{x}_{ti})\phi(\mathbf{x}_{tj})^T]$ and $\mathbf{K}_{st} = [\phi(\mathbf{x}_{si})\phi(\mathbf{x}_{tj})^T]$ refer to source domain kernel, target domain kernel and cross domain kernel respectively. The MMD function in matrix form can be represented as

$$D(\mathbf{X}_s, \mathbf{X}_t) = trace(\mathbf{KL}) \tag{3}$$

We denote L_{ij} as the element of matrix $\mathbf{L}$ in the ith row and jth column, where $\mathrm{L}_{ij} = \frac{1}{n_1^2}$ if $i \leq n_1, j \leq n_1$, $\mathrm{L}_{ij} = \frac{1}{n_2^2}$ if $n_1 < i \leq n_1 + n_2, n_1 < j \leq n_1 + n_2$, otherwise, $\mathrm{L}_{ij} = -\frac{1}{n_1 n_2}$.

B. Outlier Removal

Directly adopting domain adaptation in anti-spoofing may not achieve the desired performance because both training and testing face samples are contaminated with outlier samples. Specifically, the outlier samples can be caused by the failure of face detection, over saturation of the medium, unexpected face image blur, etc. There are two main drawbacks of involving the outlier samples in source domain data. First, training classifier can be sensitive to such outlier samples, since face spoofing detection relies on the distortion cues and the unexpected distortion from genuine samples can deteriorate the robustness of classifier. Second, the outliers can lead to a larger distance between two distributions and prevent us from learning robust transformation matrix to adapt all the source domain data to target domain [52]. Fig. 2 provides two examples of outliers in spoofing detection. The left one is the face sample which is over-saturated. The right one shows the frame where the failure of face detection happens with Viola-Jones algorithm. We can also observe from Fig. 1 that such outliers have a negative impact on compact distribution formation.

From this perspective, only the most informative data should be selected. In [53], the informative data selection strategy was applied in the cross-language text categorization task with a similar motivation. Here we focus on the facial images and remove the outlier samples from the source domain by minimizing the MMD distance before conducting domain adaptation. We transfer the idea kernel mean matching (KMM) [51] which was proposed to tackle covariate shift problem to outlier removal.

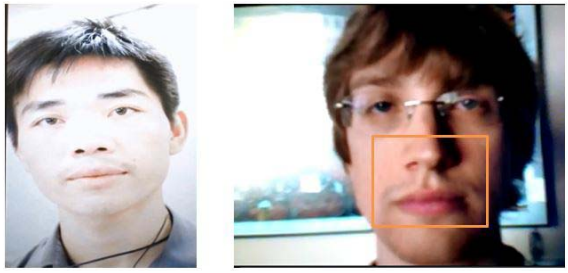

Fig. 2. Examples of outliers in CASIA and Idiap REPLAY-ATTACK databases. Left: color saturation. Right: face detection failure.

In particular, let us denote the number of inlier samples as n_1', where $n_1' < n_1$. To remove the outliers, we impose a binary correspondence matrix $\mathbf{P} \in \{0, 1\}^{n_1' \times n_1}$, which satisfies

$$\sum_{j=1}^{n_1} \mathrm{P}_{ij} = 1, \quad \forall i = 1, \ldots, n_1'$$
$$\sum_{i=1}^{n_1'} \mathrm{P}_{ij} \leq 1, \quad \forall j = 1, \ldots, n_1$$

where P_{ij} is the (i, j)th element in $\mathbf{P}$. In other words, each row of the matrix $\mathbf{P}$ indicates one and only one inlier sample from the source domain. We also introduce a vector $\mathbf{p} = \mathbf{P}^T \mathbf{1}_{n_1'}$, where $\mathbf{1}_{n_1'} \in \mathbb{R}^{n_1'}$ denotes the vectors with all ones. Since each column of $\mathbf{P} \in \{0, 1\}^{n_1' \times n_1}$ contains at most one non-zero element, we have $\mathbf{p} \in \{0, 1\}^{n_1}$ and $\sum_{i=1}^{n_1} \mathrm{p}_i = n_1'$. In other words, $\mathrm{p}_i = 1$ indicates the ith sample from $\mathbf{p}$ is selected as an inlier training sample, and $\mathrm{p}_i = 0$ indicates an outlier sample. The kernel matrix defined on all the samples after removing the outliers is then represented as

$$\mathbf{K}^* = \begin{pmatrix} \mathbf{P}\mathbf{K}_{ss}\mathbf{P}^T & \mathbf{P}\mathbf{K}_{st} \\ \mathbf{K}_{ts}\mathbf{P}^T & \mathbf{K}_{tt} \end{pmatrix} \tag{4}$$

where $\mathbf{K}^* \in \mathbb{R}^{(n_1'+n_2)\times(n_1'+n_2)}$.

To maximize the similarity of the distribution of the selected source domain samples and target domain samples, we choose $\mathbf{P}$ such that the MMD function below is minimized,

$$\min_{\mathbf{P}} \ trace(\mathbf{K}^*\mathbf{L}^*) \tag{5}$$

where $\mathbf{L}^*$ is defined as $\mathrm{L}_{ij}^* = \frac{1}{(n_1')^2}$ if $i \leq n_1', j \leq n_1'$, $\mathrm{L}_{ij}^* = \frac{1}{n_2^2}$ if $n_1' < i \leq n_1' + n_2, n_1' < j \leq n_1' + n_2$, otherwise, $\mathrm{L}_{ij}^* = -\frac{1}{n_1' n_2}$.

Then, the objective function of minimizing $\mathbf{P}$ can be rewritten as

$$\min_{\mathbf{P}} \ trace(\frac{1}{(n_1')^2}\mathbf{P}\mathbf{K}_{ss}\mathbf{P}^T\mathbf{1}_{n_1'}\mathbf{1}_{n_1'}^T - \frac{1}{n_1' n_2}\mathbf{P}\mathbf{K}_{st}\mathbf{1}_{n_2}\mathbf{1}_{n_1'}^T)$$
$$+ \ trace(\frac{1}{n_2^2}\mathbf{K}_{tt}\mathbf{1}_{n_2}\mathbf{1}_{n_2}^T - \frac{1}{n_1' n_2}\mathbf{K}_{ts}\mathbf{P}\mathbf{1}_{n_1'}\mathbf{1}_{n_2}^T) \tag{6}$$

where $\mathbf{1}_{n_1'} \in \mathbb{R}^{n_1'}$ and $\mathbf{1}_{n_2} \in \mathbb{R}^{n_2}$ denote the vectors with all ones. The above objective function can be further simplified given as,

$$\min_{\mathbf{p}} \ trace(\frac{1}{(n_1')^2}\mathbf{P}\mathbf{K}_{ss}\mathbf{P}^T\mathbf{1}_{n_1'}\mathbf{1}_{n_1'}^T - \frac{2}{n_1'n_2}\mathbf{P}\mathbf{K}_{st}\mathbf{1}_{n_2}\mathbf{1}_{n_1'}^T) \quad (7)$$

By observing that $\frac{1}{(n_1')^2}\mathbf{P}\mathbf{K}_{ss}\mathbf{P}^T\mathbf{1}_{n_1'}\mathbf{1}_{n_1'}^T = \frac{1}{(n_1')^2}\mathbf{1}_{n_1'}^T\mathbf{P}\mathbf{K}_{ss}\mathbf{P}^T\mathbf{1}_{n_1'}$, the above objective function can be further rewritten as,

$$\min_{\mathbf{p}} \ \frac{1}{(n_1')^2}\mathbf{p}^T\mathbf{K}_{ss}\mathbf{p} - \frac{2}{n_1'n_2}\mathbf{p}^T\mathbf{K}_{st}\mathbf{1}_{n_2} \quad (8)$$

This objective function is difficult to solve due to the binary constraints on $\mathbf{p}$. To get an efficient solution, we relax $\mathbf{p} \in \mathbb{R}^{n_1}$ which satieties $0 \leq \mathrm{p}_i \leq 1$ and $\sum_{i=1}^{n_1}\mathrm{p}_i = n_1'$. Therefore, p_i can be treated as a weight for the ith sample, and if p_i is smaller than a threshold, the ith sample can be treated as an outlier. The above problem is essentially a quadratic programming problem, which can be efficiently solved [54]. After obtaining the vector $\mathbf{p}$, a fixed number of training samples with lowest values are treated as outliers. In our work, for each database with different features, we first conduct cross-validations by using libSVM [55] and get an accuracy (acc%). Then, (1-acc)% of the total samples with the smallest weight values are removed as outliers. To our best knowledge, it is generally nontrivial to determine the threshold for outlier samples, and our simple approach works quite well.

C. Domain Adaptation

The main objective of domain adaptation for face spoofing detection is to model the distributions of face samples and learn a mapping function which can align the distributions from the source domain data to the target domain data. This allows us to transfer the prior knowledge of the source domain to target domain for anti-spoofing purpose. Considering this, we introduce the following two principles to further analyze this problem.

- The performance of face spoofing detection depends heavily on the distribution of facial appearances, illumination conditions and camera quality of given face samples.
- The extracted features which account for facial appearances, illumination conditions and camera quality can be approximated in a low-dimensional linear subspace.

The first observation is based on the recent studies of face spoofing detection on various databases. It can be observed that the performance of face spoofing detection by using the same features can be diverse, implying that the facial appearance, illumination condition, and adopted camera models have a significant impact on the detection accuracy. Moreover, straightforwardly imposing features for the cross-database scenario cannot achieve satisfied performance, the reason of which is mainly due to the mismatch of face capturing conditions.

The second principle has been analyzed in many face recognition tasks. By projecting facial features on eigenvectors, we can extract more robust facial information on a low-dimension manifold [57], [58]. On the other hand, the diverse reflection on the facial surface caused by different illumination conditions can also be modeled as low-dimensional linear subspace which can further improve face recognition performance [59]. Furthermore, both of the camera models and recapturing process are associated with image quality. We also observe that the influence of camera models can be dominant [60]. Therefore, it is reasonable to assume that the diverse of camera models can be formulated in low-dimension subspace.

Based on the above principles, we formulate the mapping function ϕ based on Principle Component Analysis (PCA) which is given as

$$\begin{aligned} \phi : \hat{\mathbf{X}}_s \Rightarrow \mathbf{U}_s, \quad \mathbf{X}_t \Rightarrow \mathbf{U}_t \\ \mathbf{U}_s &= \arg\max\nolimits_{\|\mathbf{U}^T\mathbf{U}\|=\boldsymbol{I}_\mathbf{d}}\{\mathbf{U}^T\hat{\mathbf{X}}_s^T\hat{\mathbf{X}}_s\mathbf{U}\} \\ \mathbf{U}_t &= \arg\max\nolimits_{\|\mathbf{U}^T\mathbf{U}\|=\boldsymbol{I}_\mathbf{d}}\{\mathbf{U}^T\mathbf{X}_t^T\mathbf{X}_t\mathbf{U}\} \end{aligned} \quad (9)$$

where $\mathbf{U}_s$ and $\mathbf{U}_t$ are the eigenspaces of source domain features after outlier removal $\hat{\mathbf{X}}_s$ and target domain features $\mathbf{X}_t$ respectively. $\boldsymbol{I}_d$ is an identity matrix with dimension d obtained via PCA.

Then the MMD function based on subspace of extracted features can be formulated as

$$D(\hat{\mathbf{X}}_s, \mathbf{X}_t) = \|\mathbf{U}_s - \mathbf{U}_t\|_F^2 \quad (10)$$

Our goal is to learn a mapping function $\mathbf{M}$ from $\mathbf{U}_s$ to $\mathbf{U}_t$ which minimizes the MMD function defined in (10). Therefore, we employ the subspace alignment (SA) algorithm [61] to minimize (10) by learning a transformation matrix from the subspace of source domain to the subspace of target domain. After obtaining the transformation matrix, the source domain data are mapped to another space based on transformation matrix and the classifier is trained via transformed source domain data. The procedures of subspace alignment is summarized as follows,

- Normalizing the source and target data by Z-Score [62].
- Applying Principle Component Analysis (PCA) to derive d eigenvectors corresponding to the top d largest eigenvalues (following the theory proposed in [61]), which are used as bases of source and target subspaces as $\mathbf{U}_s$ and $\mathbf{U}_t$.
- Learning the mapping function $\mathbf{M}$ by

$$\mathbf{M}^* = \mathrm{argmin}_{\mathbf{M}} \ \|\mathbf{U}_s\mathbf{M} - \mathbf{U}_t\|_F^2 \quad (11)$$

To extract the non-linear information induced by the subspace, we also extend the PCA to kernel PCA case by using a Gaussian kernel. We refer the kernel based method as kernel Subspace Alignment (KSA) [63]. The only difference between Subspace Alignment and kernel Subspace Alignment is that the eigenvectors in step (2) are obtained by kernel PCA on $\hat{\mathbf{X}}_s$ and $\mathbf{X}_t$.

D. Classifier Training

We seek for the representer theory [56] to train a binary classifier which can be represented as

$$y = \sum_{i=1}^{n_1'}\alpha_i y_i k(\mathbf{x}_i, \mathbf{x}) \quad (12)$$

where n_1' is the number of training samples after outlier removal, $\mathbf{x}_i$ is the training samples from $\hat{\mathbf{X}}_s$, $\mathbf{x}$ is the sample for testing from $\mathbf{X}_t$, y is the predicted score and α_i is the dual variable which can be solved by

$$\max_{\alpha_i} \ \sum_{i=1}^{n_1'}\alpha_i - \frac{1}{2}\sum_{i,j=1}^{n_1'} y_i\alpha_i k(\mathbf{x}_i, \mathbf{x}_j)y_j\alpha_j \quad (13)$$

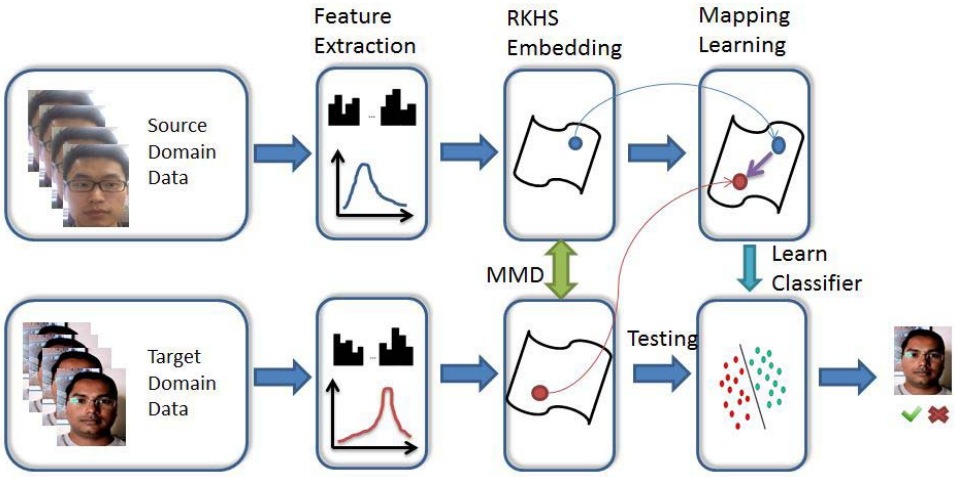

Fig. 3. The framework of face spoofing detection with unsupervised domain adaptation. After feature extraction from facial image, we first adopt a feature embedding by using Principle Component Analysis (PCA). Subsequently, we learn a mapping function to align the eigenspaces between source domain data and target domain data. Finally, a classifier is trained based on representer theory [56].

where y_i is the label of $\mathbf{x}_i$. We then introduce the kernel representation induced by the mapping function of unsupervised domain adaptation.

The optimal $\mathbf{M}^*$ of SA is obtained as $\mathbf{M}^* = \mathbf{U}_s^T\mathbf{U}_t$. Recall that the feature representation after PCA is given as $\hat{\mathbf{X}}_s\mathbf{U}_s$ and $\mathbf{X}_t\mathbf{U}_t$. Therefore, after obtaining $\mathbf{M}^*$, we can compute training kernel and testing kernel by

$$\begin{aligned}\mathbf{K}_{ss} &= \hat{\mathbf{X}}_s\mathbf{U}_s\mathbf{U}_s^T\mathbf{U}_t\mathbf{U}_t^T\mathbf{U}_s\mathbf{U}_s^T\hat{\mathbf{X}}_s^T \\ \mathbf{K}_{st} &= \hat{\mathbf{X}}_s\mathbf{U}_s\mathbf{U}_s^T\mathbf{U}_t\mathbf{U}_t^T\mathbf{X}_t^T \end{aligned} \tag{14}$$

$\mathbf{K}_{ss}$ is applied in (13) to train a binary classifier and $\mathbf{K}_{st}$ is applied for score prediction. The whole process of our algorithm is summarized in Algorithm 1. We also show the framework in Fig.3 for illustration.

Algorithm 1 Unsupervised Domain Adaptation for Face Spoofing Detection

Input : Source domain face features $\mathbf{X}_s$, target domain face features $\mathbf{X}_t$, source domain label $\mathbf{y}_s$

Output: Classifier coefficients α_i

1 Compute the kernel representation $\mathbf{K}$ based on source domain data and target domain data;
2 Optimize objective function (8) to obtain $\mathbf{p}$, then conduct outlier removal to obtain $\hat{\mathbf{X}}_s$ based on $\mathbf{p}$;
3 Learn the mapping function $\mathbf{M}^*$ based on SA or KSA (11);
4 Compute the kernel representation $\mathbf{K}_{ss}$ and $\mathbf{K}_{st}$ based on (14);
5 Train a kernel based classifier to get α_i based on (13).

IV. Feature Analysis

The state-of-the-art methods which specifically focus on developing meaningful features did not leverage the unlabeled testing samples during training. Therefore, in this paper, we do not intend to compare with the results based on features and instead employ the state-of-the-art features for this specific task. In this section, we introduce and analyze the state-of-the-art image-based features for unsupervised domain adaptation based face anti-spoofing. In particular, to ensure sufficient statistical power, both hand-crafted and convolutional neural network (CNN) based features are employed, which are all widely adopted in research community so far.

A. Hand-Crafted Features

Following the previous work on face spoofing detection [7], [9] and multimedia recapturing analysis [64], we regard the face anti-spoofing as a special image recapturing detection problem based on the observation that face spoofing is ultimately a multimedia recapturing process (we show an example in Fig.4). As such, three types of distortions are generally considered in the selection of hand-crafted features.

1) Loss of details
When a scene is captured by a physical camera device, a certain level of detail loss is introduced into the digital image. One reason originates from sharpness reduction which is caused by aberration, lens aperture distortion, color filter array demosaicing and resizing. By recapturing the printed image on a paper or displayed image on a screen, there will be a dramatic increase in the degree of loss.
2) Color distortion
The imperfection of color production originates from the imperfect camera color filter and a limited gamut of the display medium. When an image is recaptured from a display medium, there will be a significant loss of color details in the recaptured one compared with the originally captured version. The color of recaptured images tends to be distorted with different brightness, saturation [65] and contrast [66] compared to the original image.
3) Additive texture patterns
During image recapturing process, the patterns caused by the superposition of two grid structures in fine scale are also inevitable due to many reasons, e.g. low resolution of printing and more pattern artifacts of displaying devices, regular texture patterns appearing on the printed papers and screens, and unique polarity inverse driving pattern. Therefore, the additive texture patterns can also provide us useful information for spoofing detection.

(a) (b)

Fig. 4. Comparison between a genuine face and fake face. (a) Genuine face sample. (b) Fake face sample.

The pioneer work regarding face spoofing detection [19] used Fourier Spectrum, which can be viewed as the measurement of detail since the degree of loss will influence the energy of high frequency component in Fourier Spectrum. Cao and Kot [64] imposed multi-scale Discrete Wavelet Transform (DWT) for loss-of-the-detail analysis. We adopt the statistical moments from [64] as part of feature based on the loss-of-the-detail artifact where the mean and the standard deviation of the absolute subband coefficients of DWT are computed. More specifically, considering the wavelet coefficients $\{y_{Q,1}, y_{Q,2}, \ldots, y_{Q,N_Q}\}$ at the Qth subband, where N_Q is the number of coefficients in Qth subband. The mean m_Q and deviation σ_Q of the absolute coefficients are adopted as features

$$m_Q = \frac{1}{N_Q}\sum_{i=1}^{N_Q}|y_{Q,i}| \tag{15}$$

$$\sigma_Q = \sqrt{\frac{1}{N_Q}\sum_{i=1}^{N_Q}(|y_{Q,i}| - m_Q)^2} \tag{16}$$

The loss-of-the-detail artifacts are also highly correlated with image quality [67] where the spectrum coefficients are modeled in Gaussian General Density (GGD). In particular, it is defined as

$$f(x;\mu,\sigma,\gamma) = \left(\frac{\gamma}{2\sigma\Gamma(\frac{1}{\gamma})\beta}\right)e^{-(\frac{|x-\mu|}{\sigma\beta})^{\gamma}} \tag{17}$$

where μ is the mean, σ is the standard deviation, γ is the shape parameter and β is the scale parameter defined as

$$\beta = \sqrt{\frac{\Gamma(\frac{1}{\gamma})}{\Gamma(\frac{3}{\gamma})}} \tag{18}$$

where $\Gamma(\cdot)$ denotes the gamma function.

By modeling the subband coefficients with GGD, we are more interested in the shape parameter γ which controls the distribution shape. In particular, a small γ value corresponds to fewer variations of wavelet subband coefficients and vice versa. Therefore, γ can be viewed as an indicator of the amount of detail loss in the facial image. We combine the moment m_Q, σ_Q and the γ value as our feature based on the loss-of-the-detail artifact. More specifically, we conduct a 3-level DWT on R, G, and B planes respectively and extract the wavelet features.

Hand-crafted texture features (e.g. LBP) are also proved to be effective for face spoofing detection [21]. One recent work [9] conducted analyses on various types of texture features (LBP, CoALBP [68], LPQ [69], BSIF [70] and SID [71]) and found that CoALBP and LPQ can achieve the state-of-the-art performance on these public databases. Moreover, the methods are extended for color information extraction in [9] since color distortion is also an important cue for face spoofing detection. In particular, color space conversion was conducted and the texture features were extracted from three different color channels regarding the color space. The authors also showed that color space conversion can improve the cross-database detection capability. In this paper, we adopt the texture features CoALBP and LPQ as the descriptor to analyze the domain adaptation performance under the cross-database scenario. We follow [9] to extract CoALBP feature with the radius $R = \{1, 2, 4\}$ and the corresponding direction distance $B = \{2, 4, 8\}$, and LPQ feature by setting the parameters as $a = \frac{1}{7}$ and the neighboring block size as 7.

B. Deep Learning Based Feature

We train a face spoofing detection CNN based on the architecture AlexNet [38] with both genuine and fake face images. In AlexNet, there are five convolutional layers (with pooling layer and rectifier) and three fully-connected layers. The final layer is used for classification. Pre-trained AlexNet with ImageNet database is proven to be effective for fingerprint spoofing detection [34]. The recent work [72] on deep learning showed that the hidden layers transit from general to specific, implying that the shallow layers contain more general information (e.g. edge) while the deep layers tend to be customized towards a specific task. It is worth mentioning that the ImageNet ILSVRC competition is based on object recognition related tasks while the face spoofing detection mainly focuses on distortion cues which are different from object recognition. Therefore, we fix the five convolutional layers as the weights pre-trained by ImageNet which are more related to general tasks. For fully-connected layers, we change the final layer from 1000 nodes (in AlexNet) to 2 nodes since we only have 2 classes (genuine, fake) for classification. We then randomly initialize the weight of fully connected layer as Gaussian distribution $N(0, 0.001)$. During training, we back-propagate the gradient through the whole network to update the parameters. Stochastic Gradient Descent (SGD) is employed for training. We set the momentum to 0.9, learning rate to 0.001, weight decay to 0.0005 and learning rate decay to 10^{-7}.

V. Experiments and Discussions

A. Databases

For face anti-spoofing, the commonly used public databases are Idiap REPLAY-ATTACK [73], CASIA Face AntiSpoofing [3] and MSU mobile face spoofing database [7].

The Idiap REPLAY-ATTACK database [73] consists of 1200 videos taken by the webcam on a MacBook with the resolution 320 × 240. The videos were captured under two conditions: 1) the controlled condition with a uniform background and lighting, 2) the adverse condition with the complex background and natural lighting. Spoofing attack was launched by using Canon PowerShot to capture face video and the high-resolution videos were displayed using iPad 1 (1024 × 768), iPhone 3GS (480 × 320) and paper as the spoofing medium.

The CASIA Face AntiSpoofing Database [3] consists of 600 videos. Compared with Idiap REPLAY-ATTACK database, CASIA uses more face acquisition devices with different quality levels (Sony NEX-5 with the resolution 1280 × 720, two different USB cameras with the resolution 640 × 480). The spoofing types include warping attack, cutting attack and replaying attack.

The MSU mobile face spoofing database [7] consists of 280 videos of genuine and fake faces. The face videos are captured by Laptop camera and Android phone camera with resolutions of 640 × 480 and 720 × 480 respectively. The MSU database contains mainly two different spoofing attacks, printed photo attack and replay video attack. The MSU database is also divided into training and testing subsets based on different subjects. Another database, MSU unconstrained smartphone spoof attack database [12], contains 10K photos with more than 1000 subjects which are collected from the website with diverse background and illumination conditions. We use the former database in this work since we are focusing on the compact domain for domain adaptation.

This inspires us to build a new face anti-spoofing database which contains mobile phones with both high, middle and low-quality levels. In particular, important factors that need to be taken into consideration include:

- The face samples contained in the database should be captured in real-world scenarios and the spoofing medium will be able to bypass the face verification system without anti-spoofing detection.
- Face videos captured with a large variety of cameras under different quality levels and illumination settings are preferred.
- Different spoofing attacking types are better to be considered to improve the robustness of anti-spoofing.

We introduce a new and more comprehensive face anti-spoofing database, Rose-Youtu Face Liveness Detection Database, which covers a large variety of illumination conditions, camera models, and attack types. The Rose-Youtu Face Liveness Detection Database (Rose-Youtu) consists of more than 4000 videos with 25 subjects.[2] The scale is large in contrast to CASIA's 600 videos, Idiap's 1200 videos, and MSU's 280 videos. For each subject, there are 150-200 video clips with the average duration around 10 seconds. Five mobile phones were used to collect the database: (a) Hasee smartphone (with the resolution 640×480), (b) Huawei Smart-phone (with the resolution 640 × 480), (c) iPad 4 (with the resolution 640 × 480), (d) iPhone 5s (with the resolution 1280 × 720) and (e) ZTE smart-phone (with the resolution 1280 × 720). All face videos are captured by a front-facing camera. The standoff distance between face and camera is about 30 – 50 cm.

We consider three spoofing attack types including printed paper attack, video replay attack, and masking attack. For printed paper attack, face image with still printed paper and quivering printed paper (A4 size) are used. For video replay attack, we display a face video on Lenovo LCD screen (with the resolution 4096 × 2160) and Mac screen (with the resolution 2560 × 1600). For masking attack, masks with and without cropping are considered. Moreover, the face videos are captured with different backgrounds which guarantee the face videos are coupled with different illumination conditions.

[2]20 among 25 subjects are public released. The database can be downloaded through the link: http://rose1.ntu.edu.sg/Datasets/faceLivenessDetection.asp

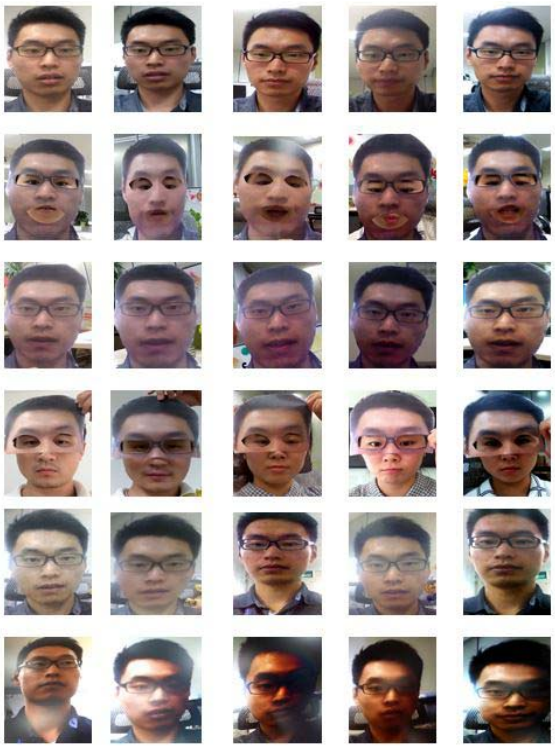

Fig. 5. Examples of Rose-Youtu Liveness Database. From top to bottom: face images in genuine, cropped mask, full mask, upper mask, paper print and video replay versions. (For paper print attack, both warped paper and still paper attacks are considered.) From left to right: face images captured by iPhone 5S, Hasee mobile phone, Huawei mobile phone, iPad and ZTE mobile phone.

To keep consistent with the genuine face video, the standoff distance between spoofing medium and camera is also about 30 – 50 cm. Some examples of our database are shown in Fig. 5. We divide the Rose-Youtu Database into training and testing subsets. Videos belonging to the first 10 indexed subjects are used for training and the others are for testing.

B. Experimental Settings

In this paper, we adopt Viola-Jones algorithm [74] for face detection. For texture feature (CoALBP, LPQ) extraction, we follow [9] to resize the face region into 64 × 64. For deep learning based feature extraction, we normalize the facial region to be 227 × 227. The original facial image is used for wavelet feature since image resizing may introduce unexpected detail loss. For cross-database scenario, we conduct domain adaptation between training and testing samples before classification. The parameters C for SVM classifier with the linear kernel is determined based on cross-validation. The PCA dimension of domain adaptation is determined by the theory in [61]. In particular, we set the deviation value equaling to 10^5 and confidence equaling to 0.1 as suggested in [61].

As mentioned before, in this paper, the state-of-the-art features are employed in the face anti-spoofing with domain adaptation. As such, the comparisons of the experimental results are conducted for each adopted feature separately, which can better reflect the performance of the proposed scheme. In the experiment, we thus evaluate the performance by considering both hand-crafted and deep learning based features to show the effectiveness our proposed method. We first show the performance of intra-database. The evaluation protocols of

TABLE I
RESULTS OF INTRA-DATABASE PERFORMANCE

	Wavelet [63]	**CoALBP (HSV)** [9]	**CoALBP (YCbCr)** [9]	**LPQ (HSV)** [9]	**LPQ (YCbCr)** [9]	**Deep Learning based Feature** [37]
CASIA	10.9%	5.5%	10.0%	7.4%	16.2%	7.6%
Idiap	9.9%	3.7%	1.4%	7.9%	6.3%	2.1%
MSU	12.8%	9.8%	8.1%	12.2%	7.4%	5.8%
Rose-Youtu	26.6%	16.4%	17.1%	30.4%	27.6%	8.0%

† Half Total Error Rate (HTER) is employed for Idiap REPLAY-ATTACK database, and Equal Error Rate is used for CASIA, MSU and Rose-Youtu databases. The results of Rose-Youtu are obtained with 25 subjects.

Idiap REPLAY-ATTACK, CASIA and MSU databases are consistent with the prior works [3], [7], [73]. In particular, Half Total Error Rate (HTER) is employed for Idiap REPLAY-ATTACK database, and Equal Error Rate is used for CASIA and MSU databases. For our new developed Rose-Youtu database, we use the first 10 indexed subjects for training and the remaining subjects for testing. The results are shown in Table I, which demonstrate the effectiveness of our adopted features. We can observe that for CASIA and Idiap REPLAY-ATTACK databases, the hand-crafted feature CoALBP achieves the best performance while deep learning based feature performs better in MSU and Rose-Youtu databases.

First, we can observe that among the features adopted, wavelet features turn out to achieve worse performance compared with other features. Such results are reasonable since the motivation of using wavelet feature is to extract the loss-of-the-detail information, which is very likely to be influenced by many factors (e.g. the quality of the camera and attack mediums, the distance between person and camera, etc). This may deteriorate the discriminative capability of wavelet feature. Deep learning based features are obtained in a data-driven manner. Therefore, it is also reasonable that deep learning based features can achieve relatively better performance among all the features. For CoALBP feature, we notice that it has achieved the best performance on some of the databases which are less diverse. However, the performance drops rapidly on the database with diverse content (e.g. Rose-Youtu database). Another interesting observation is that when using the same feature in different color spaces, the performance can be different. We conjecture the reason that different color spaces are designed for different purposes, such that the extracted features can also be encoded with different discriminative information.

Furthermore, we evaluate the performance of the cross-database face spoofing detection by imposing unsupervised domain adaptation with various features. To make fair comparisons with recent works [9], [37], HTER is used for cross-database scenario evaluation. We follow the protocol defined in [37] which divides the training database into five folds. In particular, one of them is used as the development set for threshold τ determination, and the others are used for training. The final HTER performance is obtained by averaging the results. Considering that the outlier removal method originates from kernel mean matching algorithm (KMM) [51] which was proposed for distribution alignment, we consider the KMM method as one of our baselines.

C. Cross-Database Experimental Results

Three public databases (CASIA, Idiap REPLAY-ATTACK, and MSU) and our newly developed Rose-Youtu liveness database are used for cross-database spoofing detection evaluation. To facilitate the comparison, one database (composed of both training and testing folds together) is used for training and another is used for testing. Thus we have 12 scenarios in total. Various feature representations are considered for evaluation. However, deep learning based feature is only considered when using Rose-Youtu database for either training or testing since we observe that only the deep learning based feature can achieve desired performance on Rose-Youtu database. For unsupervised domain adaptation, we report the results with and without outlier removal. We use "w/o DA" to represent without domain adaptation and "§" to denote the domain adaptation with outlier removal. We also use "C", "I", "M" and "Y" to denote CASIA, Idiap REPLAY-ATTACK, MSU and Rose-Youtu database respectively. The experimental results for cross-database face spoofing detection are shown from Table II to Table VII.

TABLE II
PERFORMANCE (HTER) OF CROSS-DATABASE WITH WAVELET STATISTICAL FEATURE

	C → I	C → M	I → C	I → M	M → C	M → I
w/o DA	49.9%	49.2%	47.7%	48.6%	49.1%	50.0%
KMM	50.0%	43.7%	51.1%	45.2%	46.2%	49.9%
SA	39.6%	36.9%	35.2%	33.1%	48.4%	37.9%
KSA	37.5%	22.1%	36.7%	35.8%	42.4%	45.6%
SA§	36.8%	28.5%	34.3%	**31.3**%	41.4%	35.2%
KSA§	**33.1**%	**19.1**%	**32.1**%	33.9%	**41.2**%	**35.1**%

† "C", "I" and "M" denote CASIA, Idiap REPLAY-ATTACK and MSU database respectively, $A \rightarrow B$ refers to using database A for training and B for testing, "§" refers to the domain adaptation technique with outlier removal.

1) Database and Feature Analysis: Based on the results, we can observe that cross-database performance highly depends on the databases we use. For example, when using CASIA and MSU for training and testing, we can achieve around 25% HTER with CoALBP in HSV color space, around 15% HTER with CoALBP in YCbCr color space and deep learning based feature. Moreover, deep learning based feature also achieves around 30% HTER under several other scenarios. However, we can only obtain close to 50% for other cases, which correspond to the random-guess scenario. We conjecture the reason may lie in that the face videos from these two databases were captured under dissimilar illumination conditions. Moreover, the influences of illumination appear to be more dominant in face spoofing detection task compared with facial appearance and camera quality, since we cannot achieve satisfied performance based on the cross-database evaluation

TABLE III
PERFORMANCE (HTER) OF CROSS-DATABASE WITH COALBP (HSV) FEATURE

	C → I	C → M	I → C	I → M	M → C	M → I
w/o DA	50.3%	24.9%	50.0%	50.0%	50.0%	50.0%
KMM	50.0%	22.4%	54.3%	60.0%	34.9%	51.9%
SA	40.2%	22.9%	37.5%	31.5%	40.9%	35.4%
KSA	37.5%	21.6%	41.2%	32.9%	37.3%	39.0%
SA§	**33.4**%	21.7%	**33.2**%	29.2%	37.7%	**30.6**%
KSA§	35.1%	**20.9**%	39.8%	**29.0**%	**34.2**%	36.9%

† "C", "I" and "M" denote CASIA, Idiap REPLAY-ATTACK and MSU database respectively, $A \rightarrow B$ refers to using database A for training and B for testing, "§" refers to the domain adaptation technique with outlier removal.

TABLE IV
PERFORMANCE (HTER) OF CROSS-DATABASE WITH COALBP (YCBCR) FEATURE

	C → I	C → M	I → C	I → M	M → C	M → I
w/o DA	50.0%	15.1%	50.1%	50.0%	44.8%	50.0%
KMM	50.0%	15.0%	50.0%	50.0%	42.8%	51.0%
SA	46.1%	15.4%	39.3%	41.7%	35.5%	51.3%
KSA	37.5%	21.2%	41.6%	33.1%	37.3%	42.9%
SA§	45.3%	**14.9**%	**34.5**%	40.5%	34.9%	47.9%
KSA§	**35.1**%	20.9%	39.7%	**29.0**%	**34.2**%	**36.9**%

† "C", "I" and "M" denote CASIA, Idiap REPLAY-ATTACK and MSU database respectively, $A \rightarrow B$ refers to using database A for training and B for testing, "§" refers to the domain adaptation technique with outlier removal.

TABLE V
PERFORMANCE (HTER) OF CROSS-DATABASE WITH LPQ (HSV) FEATURE

	C → I	C → M	I → C	I → M	M → C	M → I
w/o DA	45.5%	54.9%	43.7%	53.5%	58.7%	59.1%
KMM	44.1%	52.8%	41.0%	57.5%	49.5%	50.7%
SA	37.5%	28.0%	41.4%	29.5%	50.0%	35.9%
KSA	38.3%	41.0%	42.7%	37.5%	42.2%	39.4%
SA§	**33.4**%	**21.2**%	39.1%	**24.9**%	42.8%	**33.2**%
KSA§	36.4%	39.7%	**39.0**%	35.7%	**39.8**%	37.8%

† "C", "I" and "M" denote CASIA, Idiap REPLAY-ATTACK and MSU database respectively, $A \rightarrow B$ refers to using database A for training and B for testing, "§" refers to the domain adaptation technique with outlier removal.

performances between Idiap REPLAY-ATTACK with these features.

2) Outlier Analysis: We then revisit the characteristics of samples which are likely to be categorized as outliers. Basically, outliers are jointly determined by the source and target domains. When CASIA database is used for source domain, the blurry and low-contrast samples are treated as outliers since the samples from other databases have relatively better quality and contrast. For Idiap Replay-Attack database, we can see that outliers samples can contain glasses reflection, which is reasonable since we notice that the reflection from other databases is not as strong as Idiap Replay-Attack. For MSU database, we observe that the illumination plays an important role. (For other databases there are not many strong illumination samples.) For Rose-Youtu database, we notice that all the subjects have outlier samples. We further analyze the content of the database and find the outliers may come from the unexpected motion (due to camera moving), the distance between camera and client, and blurring. We list several outlier samples in Fig. 6.

TABLE VI
PERFORMANCE (HTER) OF CROSS-DATABASE WITH LPQ (YCBCR) FEATURE

	C → I	C → M	I → C	I → M	M → C	M → I
w/o DA	43.9%	44.3%	49.9%	46.2%	46.8%	50.0%
KMM	48.4%	49.4%	50.0%	36.6%	48.5%	49.7%
SA	47.9%	25.2%	39.4%	30.2%	36.5%	35.8%
KSA	43.1%	25.3%	44.6%	38.1%	46.0%	46.5%
SA§	**40.7**%	16.9%	**33.1**%	**27.8**%	**33.3**%	**31.8**%
KSA§	41.2%	**16.3**%	42.0%	32,5%	43.5%	41.6%

† "C", "I" and "M" denote CASIA, Idiap REPLAY-ATTACK and MSU database respectively, $A \rightarrow B$ refers to using database A for training and B for testing, "§" refers to the domain adaptation technique with outlier removal.

Fig. 6. Outlier samples selected by the proposed method. The first row shows the genuine samples and the second row shows the fake samples. The samples are collected from CASIA, Idiap REPLAY-ATTACK, MSU and Rose-Youtu databases (from left to right).

3) Benefits of Features and Domain Adaptation: Based on the results from the cross-database scenario, we can find that our proposed unsupervised domain adaptation scheme can significantly improve the performance of cross-database face spoofing detection, which indicates the effectiveness of domain adaptation scheme. We also show the bar figure in Fig. 7 by comparing the best performance at each domain adaptation scenario for different hand-crafted features. We notice that the results are variant among different features. This observation shows that the powerful features across domain and the unsupervised domain adaptation technique jointly improve the spoofing detection performance. This further provides useful evidence that our framework is effective in terms of the generalization ability of face spoofing detection. However, it is also worth mentioning that the bottleneck of domain adaptation exists when domain adaptation only achieves little performance improvement. One example can be found by using CASIA for training and MSU for testing.

TABLE VII
PERFORMANCE (HTER) OF CROSS-DATABASE WITH DEEP LEARNING BASED FEATURE

	C → I	C → M	C → Y	I → C	I → M	I → Y	M → C	M → I	M → Y	Y → C	Y → I	Y → M
w/o DA	45.8%	15.6%	46.8%	34.4%	68.6%	48.0%	50.1%	49.9%	31.0%	32.6%	43.6%	28.4%
KMM	44.9%	14.1%	46.7%	25.9%	49.7%	47.9%	32.1%	51.2%	31.3%	31.6%	43.6%	27.8%
SA	41.4%	**14.0%**	32.7%	36.5%	35.4%	43.2%	13.4%	34.9%	30.3%	35.0%	38.5%	29.0%
KSA	43.1%	16.9%	34.1%	22.0%	39.1%	40.4%	19.8%	37.2%	30.8%	33.9%	42.0%	29.2%
SA§	**39.2%**	14.3%	**31.6%**	26.3%	**33.2%**	42.8%	10.1%	**33.3%**	**30.0%**	30.7%	**36.2%**	**24.9%**
KSA§	39.3%	15.1%	33.9%	**12.3%**	33.3%	**40.1%**	**9.1%**	34.9%	30.4%	**30.1%**	38.8%	26.1%

† "C", "I", "M" and "Y" denote CASIA, Idiap REPLAY-ATTACK, MSU and Rose-Youtu database respectively, $A \rightarrow B$ refers to using database A for training and B for testing, "§" refers to the domain adaptation technique with outlier removal.

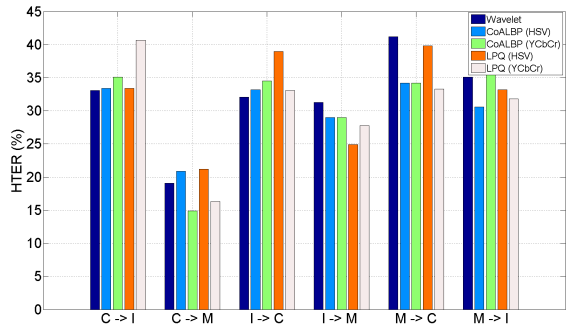

Fig. 7. The best performance in each domain adaptation scenario for each feature type.

We can observe that already a relatively good cross-database performance with CoALBP and deep learning based features can be achieved, which implies that the feature distributions of these two domains could be similar. Therefore, domain adaptation may not drive further performance improvement in this scenario. Moreover, we can observe the asymmetric performance by reversing the database for training and testing. For example, significant performance improvement can be achieved from MSU to CASIA with CoALBP feature and deep learning based feature. However, a little performance gain is observed for the scenario of CASIA to MSU. We conjecture two possible reasons regarding this issue.

- The performance can be highly relevant with the employed databases. Since CASIA database is considered to be more diverse compared with MSU due to the variations of camera models and attacking types, it is reasonable that training on CASIA can generalize better compared with training on MSU. Therefore, we observe a large performance gap between training on CASIA and training on MSU. After conducting domain adaptation, we observe that such performance gap drops significantly, and the results of training on CASIA and testing on MSU are much closer to the results training on MSU and testing on CASIA (compared with the original 15.6% and 50.1%), which demonstrate the effectiveness of domain adaptation that aligns two different distributions closer. We can also observe that the performance is only improved by 0.5%, by training on CASIA and testing on MSU. The reason for this phenomenon may lie in that the baseline model has already achieved a relatively good performance.
- On the other hand, although subspace alignment learns a transformation (mapping) from one domain to another, the classifiers trained by source domain are different and this can also lead to different performances. Moreover, we also observe that although the improvement gaps are different for different databases, our framework can generally improve the accuracy based on different types of features and databases.

We also notice that domain adaptation may not achieve significant performance improvements in some other cases (e.g. training on MSU and testing on Rose-Youtu). In essence, domain adaptation aims to solve the classification problem when the training samples and testing samples have different distributions. However, as indicated in [75], domain adaptation cannot fully address the problem when there is a dramatic deviation for the distribution of target domain from source domain encoded by a specific feature. For the Rose-Youtu database, the face samples are captured in an uncontrolled condition by various camera models under diverse illumination conditions. As such, we conjecture that the distribution of the samples from Rose-Youtu database is quite different from other databases due to the uncontrolled capturing conditions of Rose-Youtu database. Such large gap cannot be compensated by domain adaptation either due to the inherent limitation of the domain adaptation [75]. Therefore, we notice that by using Rose-Youtu database for both training and testing with domain adaptation, the performance does not have significant improvement for most of the cases.

Another interesting observation is that outlier removal can further boost the domain adaptation performances with at least $2-3\%$. This is reasonable since removing the outliers can provide us reliable samples for conducting domain adaptation and learning the powerful prediction model, as shown in the feature visualization in Fig. 1. The asymmetric performance can also be observed with/without outlier removal when considering domain adaptation, as little improvement can be achieved when training on CASIA and testing on MSU while large improvement can be obtained when using MSU for training and CASIA for testing. In addition to the reasons aforementioned, another explanation for this phenomenon lies in that the percentages of outliers can also be different by using different types of features even when we conduct outlier removal on the same database. The average percentages of removal outlier samples are shown in Table VIII. Therefore, by using different types of features, the numbers of outlier samples are different. This may result in the asymmetric

TABLE VIII

PERCENTAGE OF OUTLIER SAMPLES FOR DIFFERENT DATABASES

	C	I	M	Y
Multi-scale Wavelet	6.5%	6.2%	11.5%	–
CoALBP (YCbCr)	6.5%	3.8%	6.4%	–
CoALBP (HSV)	6.3%	8.8%	12.1%	–
LPQ (YCbCr)	9.8%	10.2%	6.7%	–
LPQ (HSV)	10.6%	11.6%	10.0%	–
Deep Learning	4.4%	2.4%	3.5%	6.7%

TABLE IX

PERFORMANCE (HTER) COMPARISONS BETWEEN OR+DA AND DA+OR (COALBP HSV)

	C → I	C → M	I → C	I → M	M → C	M → I
w/o DA	50.3%	24.9%	50.0%	50.0%	50.0%	50.0%
only DA	40.2%	22.9%	37.5%	31.5%	40.9%	35.4%
OR+DA	33.4%	21.7%	33.2%	29.2%	37.7%	30.6%
DA+OR	37.5%	22.8%	37.6%	29.1%	34.0%	35.2%

TABLE X

PERFORMANCE (HTER) COMPARISONS BETWEEN OR+DA AND DA+OR (DEEP LEARNING BASED FEATURE)

	C → I	C → M	I → C	I → M	M → C	M → I
w/o DA	45.8%	15.6%	34.4%	68.6%	50.1%	49.9%
only DA	41.4%	14.0%	36.5%	35.4%	13.4%	34.9%
OR+DA	39.2%	14.3%	26.3%	33.2%	10.1%	33.3%
DA+OR	41.0%	13.9%	33.7%	34.9%	13.4%	33.3%

performances as well. In addition, since in some scenarios, the model has already achieved relatively good performance, both domain adaptation and outlier removal will only achieve limited improvement.

Moreover, we swap the process of domain adaptation and outlier removal to analyze the influences of outlier removal process on classifier training. The results of CoALBP feature in HSV space and deep learning based feature by using Subspace Alignment are listed in Tables IX and X. It is observed that outlier removal process is effective for domain adaptation. Moreover, outlier removal can also help with training a better classifier by comparing the results of only domain adaptation with the results of first performing domain adaptation followed by outlier removal.

4) Influences of Different Number of Samples: We further analyze the influences of the number of data from the target domain on the final performance with SA method. Specifically, we conduct an experiment with CoALBP feature in HSV space by using different percentages of samples from the target domain. The results are shown in Fig. 8, where the performance with different percentage of target domain samples (1%, 5%, 10%, 20%, 50%, 100%) are demonstrated. It is also worth mentioning that the results are obtained by averaging the results for five times. We can see that the more target domain samples we can get, the lower HTER can be achieved.

5) Unbalanced Domain Adaptation: In practice, it may not be convenient to collect fake face data. To investigate the scenario that only genuine samples are available for domain adaptation in the target domain, we conduct experiments by using both hand-crafted feature (CoALBP feature on HSV space) and deep learning based feature by KSA method since we observe KSA can achieve better performance when only obtaining genuine data in hand. The results are shown in Tables XI&XII. Based on the results, we find that the

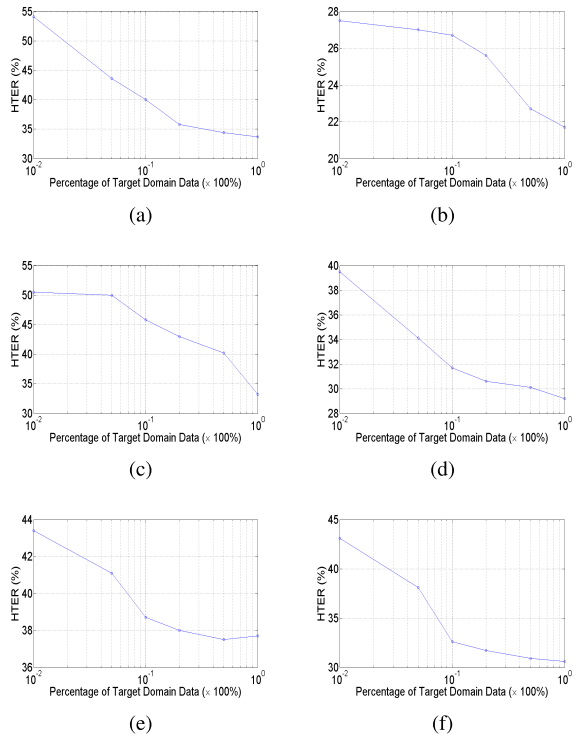

Fig. 8. HTER performance (with SA method) with the variation of target domain data number. (a) train: CASIA, test: Idiap, (b) train: CASIA, test: MSU, (c) train: Idiap, test: CASIA, (d) (train: Idiap, test: MSU), (e) train: MSU, test: CASIA, (f) train: MSU, test: Idiap

performance drop is marginal when only using genuine samples for domain adaptation for most of the cases. Moreover, in almost all cases, the domain adaptation with genuine samples perform much better compared with the method without domain adaptation. This can be explained by the reason that only genuine samples can still provide valuable cross-domain information (e.g. facial appearance, lighting, camera quality) for domain adaptation.

However, for some databases the performance drops a lot when only adopting only genuine samples compared with the results by adopting both genuine and fake samples. This may originate from the distinct characteristics of the databases. In some databases (e.g. Idiap and MSU), the illumination condition and motion information are more consistent for genuine and fake samples. As such, the distribution of the target domain will not change significantly after we remove the fake samples. However, it can be noticed that in some other databases (e.g. CASIA and Rose-Youtu) large motion variations and diverse illumination conditions exist, especially when comparing genuine and fake samples. In this case, the distribution between the source and target domain will be changed, and this may significantly influence the effectiveness of domain adaptation. Moreover, when the performance without domain adaptation can already reach a saturation level, domain adaptation may not help a lot no matter what data are finally used in target domain.

To analyze the influence by using the different number of genuine data, we also conduct an experiment with CoALBP feature in HSV space by using different percentage (1%, 5%, 10%, 20%, 50%, 100%) of genuine samples from the target domain. The results are shown in Fig. 9. We can observe that the more target domain samples we can obtain, the lower

TABLE XI

HTER RESULTS (%) BY KERNEL SUBSPACE ALIGNMENT WITH COALBP (HSV)

	C → I	C → M	I → C	I → M	M → C	M → I
w/o DA	50.3	24.9	50.0	50.0	50.0	50.0
Genuine and Fake	35.1	20.9	39.8	29.0	34.2	36.9
Only Genuine	42.5	26.5	46.3	33.4	36.4	45.6

TABLE XII

HTER RESULTS (%) BY KERNEL SUBSPACE ALIGNMENT WITH DEEP LEARNING BASED FEATURE

	C → I	C → M	I → C	I → M	M → C	M → I
w/o DA	45.8	15.6	34.4	68.6	50.1	49.9
Genuine and Fake	39.3	15.1	12.3	33.3	9.1	34.9
Only Genuine	44.2	16.1	28.0	37.5	13.6	40.2

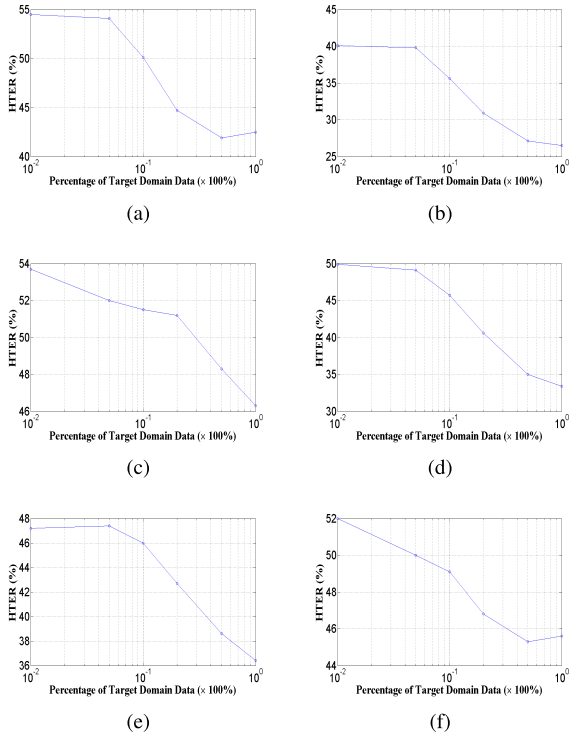

Fig. 9. HTER performance (with KSA method) with the variation of the number of genuine target domain data. (a) train: CASIA, test: Idiap, (b) train: CASIA, test: MSU, (c) train: Idiap, test: CASIA, (d) (train: Idiap, test: MSU), (e) train: MSU, test: CASIA, (f) train: MSU, test: Idiap

HTER can be achieved. This trend is consistent with the results when using both genuine and fake data for domain adaptation. This further provides useful evidence on the effectiveness of the proposed scheme.

6) Comparison With Domain Adaptation Based Feature Synthesis: In [13], a domain adaptation based feature synthesis method was proposed to tackle the cross-domain face anti-spoofing problem. Considering that it is usually applicable to collect genuine samples in the target domain, we compare the performance of the domain adaptation approach in [13] with our proposed method. Specifically, we divide our target domain into two parts. One part has labeled genuine face samples which are used for fake feature synthesis and classifier training, and the other part has no label information which is used for testing. The Center Shift algorithm reported in [13] is employed to learn the mapping function based on source domain data. The results are shown in Tables XIII and XIV.

TABLE XIII

PERFORMANCE COMPARISONS BETWEEN FEATURE SYNTHESIS [13] AND OUR METHODS. (COALBP HSV)

	C → I	C → M	I → C	I → M	M → C	M → I
w/o DA	50.3	24.9	50.0	50.0	50.0	50.0
[13]	51.2	28.9	51.3	49.1	40.2	52.2
Ours	33.4	21.7	33.2	29.2	37.7	30.6

TABLE XIV

PERFORMANCE COMPARISONS BETWEEN FEATURE SYNTHESIS [13] AND OUR METHOD. (DEEP LEARNING BASED FEATURE)

	C → I	C → M	I → C	I → M	M → C	M → I
w/o DA	45.8	15.6	34.4	68.6	50.1	49.9
[13]	49.2	18.1	39.6	36.7	49.6	49.6
Ours	39.2	14.3	26.3	33.2	10.1	33.3

As we can see from the results, such adaptation method cannot improve the performance of cross-domain face spoofing detection. This may be reasonable since our work is different from [13] in several ways as stated as follow.

- The motivation of [13] is that the fake samples of target domain are difficult to collect. Therefore, they proposed a feature synthesis method by utilizing the labeled data in the target domain. By contrast, our work is conducted in a totally unsupervised manner and no labeled information is available in the target domain.
- The work from [13] is based on prior knowledge of person identification and camera information. Therefore, a linear adaptation can be employed for feature synthesis. Their assumptions are reasonable based on the applications such as door access control,ž where person identity and camera information are available. However, we are dealing with uncontrolledž cross-domain face spoofing detection where the face identity, camera model are totally different from the training data. Their assumptions no longer hold in such scenario.

7) Results by Using Concatenated Feature: As discussed in [9], concatenating different features usually leads to a better performance than using each individual one. Therefore, we follow [9] to concatenate the CoALBP and LPQ feature in HSV and YCbCr color space into a single feature and conduct cross-database face spoofing detection experiments based on CASIA, Idiap REPLAY-ATTACK, and MSU databases. The results are reported in Table XV. We observe that all methods are generally improved by using the concatenated feature.

TABLE XV
PERFORMANCE (HTER) OF CROSS-DATABASE WITH CONCATENATED FEATURE

	C → I	C → M	I → C	I → M	M → C	M → I	Average
w/o DA (reported in [9])	30.3	20.4	37.7	34.1	46.0	33.9	33.7
w/o DA (reproduced by ourselves)	29.6	**18.5**	39.9	29.1	44.7	37.0	33.1
SA	35.8	21.7	42.4	20.8	41.2	25.4	31.2
KSA	30.5	21.0	38.5	20.9	43.0	26.8	30.1
SA§	33.9	20.2	41.4	**18.6**	40.5	**23.3**	29.7
KSA§	**27.4**	20.3	**36.0**	**18.6**	**40.1**	24.0	**27.7**

Our proposed method can still achieve performance improvement in most of the cases, which again shows the effectiveness of our proposed domain adaptation scheme for cross-domain face spoofing detection using different types of features.

D. Discussions

Practically, deploying face anti-spoofing system in real application scenarios requires the algorithm to have a good generalization ability to deal with various acquisition conditions. Since the creations of these databases are totally independent, our experiments take advantage of this benefit and show that the proposed approach can significantly improve the performance in cross-database scenarios. This further demonstrates the strong generalization ability of the scheme. As the first attempt on unsupervised domain adaptation in face anti-spoofing, several limitations existing in our approach should be improved in the future.

Firstly, although state-of-the-art features are incorporated in the current framework, how to design and learn sophisticated features that can better fit the domain adaptation scheme should be further investigated. Secondly, current domain adaptation scheme requires sufficient data to transfer the knowledge from the source domain to the target domain, and in the future domain adaptation with zero-shot learning strategy will be studied. As such, we can transfer the pre-trained knowledge even to a single face image. Finally, the complexity of the proposed scheme will be further reduced by speeding up the kernel method to meet the real-time requirement of face spoofing detection.

VI. CONCLUSION

We propose a novel framework utilizing advanced unsupervised domain adaptation algorithms for face anti-spoofing. The novelty of this framework lies in transferring the feature space of face samples from the labeled source domain to the unlabeled target domain, such that reliable model can be learned for spoofing detection. State-of-the-art hand-crafted and deep learning based features are incorporated into the domain adaptation framework and their classification accuracies are further evaluated. Extensive experiments have been conducted based on the available databases and our new database. The results show that we can achieve clearly improved generalization ability with an average of 20% improvement by domain adaptation as compared with the straightforward learning approach without domain adaptation.

ACKNOWLEDGMENT

This research was carried out at the Rapid-Rich Object Search (ROSE) Lab at the Nanyang Technological University, Singapore. The ROSE Lab is supported by the National Research Foundation, Singapore, and the Infocomm Media Development Authority, Singapore. The authors would like to thank Dr. Leida Li in Nanyang Technology University, Singapore for his helpful suggestions. We would also like to thank the Associate Editor and anonymous reviewers for their valuable comments that significantly helped us in improving the quality of the paper. We are also grateful to Mr. Zinelabidine Boulkenafet for sharing the code with us.

REFERENCES

[1] N. M. Duc and B. Q. Minh, "Your face is not your password," in *Proc. Black Hat Conf.*, vol. 1. 2009, p. 158.
[2] Z. Akhtar, C. Micheloni, and G. L. Foresti, "Biometric liveness detection: Challenges and research opportunities," *IEEE Security Privacy*, vol. 13, no. 5, pp. 63–72, Sep./Oct. 2015.
[3] Z. Zhang, J. Yan, S. Liu, Z. Lei, D. Yi, and S. Z. Li, "A face antispoofing database with diverse attacks," in *Proc. IEEE Int. Conf. Biometrics (ICB)*, Mar./Apr. 2012, pp. 26–31.
[4] 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.
[5] Y. Xu, T. Price, J.-M. Frahm, and F. Monrose, "Virtual U: Defeating face liveness detection by building virtual models from your public photos," in *Proc. USENIX Secur. Symp.*, 2016, pp. 497–512.
[6] T. de Freitas Pereira, A. Anjos, J. M. De Martino, and S. Marcel, "Can face anti-spoofing countermeasures work in a real world scenario?" in *Proc. IEEE Int. Conf. Biometrics (ICB)*, Jun. 2013, pp. 1–8.
[7] 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.
[8] 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.
[9] 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.
[10] 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.
[11] K. Patel, H. Han, and A. K. Jain, "Cross-database face antispoofing with robust feature representation," in *Proc. Chin. Conf. Biometric Recognit.*, 2016, pp. 611–619.
[12] K. Patel, H. Han, and A. Jain, "Secure face unlock: Spoof detection on smartphones," *IEEE Trans. Inf. Forensics Security*, vol. 11, no. 10, pp. 2268–2283, Jun. 2016.
[13] J. Yang, Z. Lei, D. Yi, and S. Li, "Person-specific face antispoofing with subject domain adaptation," *IEEE Trans. Inf. Forensics Security*, vol. 10, no. 4, pp. 797–809, Apr. 2015.
[14] W. Bao, H. Li, N. Li, and W. Jiang, "A liveness detection method for face recognition based on optical flow field," in *Proc. Int. Conf. Image Anal. Signal Process.*, Apr. 2009, pp. 233–236.
[15] K. Kollreider, H. Fronthaler, and J. Bigun, "Evaluating liveness by face images and the structure tensor," in *Proc. Workshop Autom. Identificat. Adv. Technol.*, Oct. 2005, pp. 75–80.
[16] G. Pan, L. Sun, Z. Wu, and S. Lao, "Eyeblink-based anti-spoofing in face recognition from a generic Webcamera," in *Proc. Int. Conf. Comput. Vis.*, Oct. 2007, pp. 1–8.
[17] G. Pan, L. Sun, and Z. Wu, *Liveness Detection for Face Recognition*. Rijeka, Croatia: INTECH, 2008.

[18] A. Anjos, M. M. Chakka, and S. Marcel, "Motion-based countermeasures to photo attacks in face recognition," *IET Biometrics*, vol. 3, no. 3, pp. 147–158, Sep. 2014.

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

[20] 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—ECCV*. Berlin, Germany: Springer, 2010, pp. 504–517.

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

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

[23] J. Yang, Z. Lei, S. Liao, and S. Z. Li, "Face liveness detection with component dependent descriptor," in *Proc. Int. Conf. Biometrics (ICB)*, Jun. 2013, pp. 1–6.

[24] D. Gragnaniello, G. Poggi, C. Sansone, and L. Verdoliva, "An investigation of local descriptors for biometric spoofing detection," *IEEE Trans. Inf. Forensics Security*, vol. 10, no. 4, pp. 849–863, Apr. 2015.

[25] 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. 2016.

[26] T. de Freitas Pereira, A. Anjos, J. M. De Martino, and S. Marcel, "*LBP—TOP* based countermeasure against face spoofing attacks," in *Proc. Workshops Comput. Vis. ACCV*, 2013, pp. 121–132.

[27] 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.

[28] 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.

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

[30] R. T. Tan and K. Ikeuchi, "Separating reflection components of textured surfaces using a single image," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 27, no. 2, pp. 178–193, Feb. 2005.

[31] F. Crete, T. Dolmiere, P. Ladret, and M. Nicolas, "The blur effect: Perception and estimation with a new no-reference perceptual blur metric," *Proc. SPIE*, vol. 6492, p. 64920I, Feb. 2007.

[32] P. Marziliano, F. Dufaux, S. Winkler, and T. Ebrahimi, "A no-reference perceptual blur metric," in *Proc. Int. Conf. Image Process. (ICIP)*, vol. 3. Sep. 2002, pp. III-57–III-60.

[33] Y. Chen, Z. Li, M. Li, and W.-Y. Ma, "Automatic classification of photographs and graphics," in *Proc. IEEE Int. Conf. Multimedia Expo (ICME)*, Jul. 2006, pp. 973–976.

[34] R. F. Nogueira, R. de Alencar Lotufo, and R. C. Machado, "Fingerprint liveness detection using convolutional neural networks," *IEEE Trans. Inf. Forensics Security*, vol. 11, no. 6, pp. 1206–1213, Jun. 2016.

[35] J. Deng, W. Dong, R. Socher, L.-J. Li, K. Li, and L. Fei-Fei, "ImageNet: A large-scale hierarchical image database," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2009, pp. 248–255.

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

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

[38] A. Krizhevsky, I. Sutskever, and G. E. Hinton, "ImageNet classification with deep convolutional neural networks," in *Proc. Adv. Neural Inf. Process. Syst.*, 2012, pp. 1097–1105.

[39] 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. IEEE Int. Conf. Biometrics (ICB)*, Jun. 2013, pp. 1–6.

[40] Z. Zhang, D. Yi, Z. Lei, and S. Z. Li, "Face liveness detection by learning multispectral reflectance distributions," in *Proc. Int. Conf. Autom. Face Gesture Recognit. Workshops*, Mar. 2011, pp. 436–441.

[41] G. Chetty, "Biometric liveness checking using multimodal fuzzy fusion," in *Proc. IEEE Int. Conf. Fuzzy Syst.*, Jul. 2010, pp. 1–8.

[42] A. Sepas-Moghaddam, L. Malhadas, P. Correia, and F. Pereira, "Face spoofing detection using a light field imaging framework," *IET Biometrics*, vol. 7, no. 1, pp. 39–48, Jan. 2018.

[43] V. Conotter, E. Bodnari, G. Boato, and H. Farid, "Physiologically-based detection of computer generated faces in video," in *Proc. Int. Conf. Image Process. (ICIP)*, Oct. 2014, pp. 248–252.

[44] V. N. Vapnik and V. Vapnik, *Statistical Learning Theory*, vol. 1. New York, NY, USA: Wiley, 1998.

[45] A. Bergamo and L. Torresani, "Exploiting weakly-labeled Web images to improve object classification: A domain adaptation approach," in *Proc. Adv. Neural Inf. Process. Syst.*, 2010, pp. 181–189.

[46] W. Li, L. Duan, D. Xu, and I. W. Tsang, "Learning with augmented features for supervised and semi-supervised heterogeneous domain adaptation," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 36, no. 6, pp. 1134–1148, Jun. 2014.

[47] X. Cao, D. Wipf, F. Wen, G. Duan, and J. Sun, "A practical transfer learning algorithm for face verification," in *Proc. Int. Conf. Comput. Vis.*, Dec. 2013, pp. 3208–3215.

[48] L. van der Maaten and G. Hinton, "Visualizing data using t-SNE," *J. Mach. Learn. Res.*, vol. 9, pp. 2579–2605, Nov. 2008.

[49] A. Gretton, K. M. Borgwardt, M. J. Rasch, B. Schölkopf, and A. Smola, "A kernel two-sample test," *J. Mach. Learn. Res.*, vol. 13, no. 1, pp. 723–773, 2012.

[50] M. Long, J. Wang, G. Ding, D. Shen, and Q. Yang, "Transfer learning with graph co-regularization," *IEEE Trans. Knowl. Data Eng.*, vol. 26, no. 7, pp. 1805–1818, Jul. 2014.

[51] J. Huang, A. Gretton, K. M. Borgwardt, and B. Schölkopf, and A. J. Smola, "Correcting sample selection bias by unlabeled data," in *Proc. Adv. Neural Inf. Process. Syst.*, 2006, pp. 601–608.

[52] R. Aljundi, R. Emonet, D. Muselet, and M. Sebban, "Landmarks-based kernelized subspace alignment for unsupervised domain adaptation," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2015, pp. 56–63.

[53] J. T. Zhou, S. J. Pan, I. W. Tsang, and S.-S. Ho, "Transfer learning for cross-language text categorization through active correspondences construction," in *Proc. 13th AAAI Conf. Artif. Intell.*, 2016, pp. 2400–2406.

[54] N. Gould and P. L. Toint, "Preprocessing for quadratic programming," *Math. Program.*, vol. 100, no. 1, pp. 95–132, 2004.

[55] C.-C. Chang and C.-J. Lin, "LIBSVM: A library for support vector machines," *ACM Trans. Intell. Syst. Technol.*, vol. 2, no. 3, pp. 27:1–27:27, 2011.

[56] B. Schölkopf, R. Herbrich, and A. J. Smola, "A generalized representer theorem," in *Proc. Int. Conf. Comput. Learn. Theory*, 2001, pp. 416–426.

[57] X. Jiang, "Asymmetric principal component and discriminant analyses for pattern classification," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 31, no. 5, pp. 931–937, May 2009.

[58] X. Jiang, B. Mandal, and A. Kot, "Eigenfeature regularization and extraction in face recognition," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 30, no. 3, pp. 383–394, Mar. 2008.

[59] R. Basri and D. W. Jacobs, "Lambertian reflectance and linear subspaces," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 25, no. 2, pp. 218–233, Feb. 2003.

[60] H. Li, S. Wang, and A. C. Kot, "Face spoofing detection with image quality regression," in *Proc. IEEE Int. Conf. Image Process. Theory, Tools Appl.*, Dec. 2016, pp. 1–6.

[61] B. Fernando, A. Habrard, M. Sebban, and T. Tuytelaars, "Unsupervised visual domain adaptation using subspace alignment," in *Proc. Int. Conf. Comput. Vis.*, Dec. 2013, pp. 2960–2967.

[62] M. Rjam, *An Introduction to Mathematical Statistics and its Applications*. London, U.K.: Pearson, 2000.

[63] W. Li, L. Chen, D. Xu, and L. Van Gool, "Visual recognition in RGB images and videos by learning from RGB-D data," *IEEE Trans. Pattern Anal. Mach. Intell.*, 2017.

[64] H. Cao and A. C. Kot, "Identification of recaptured photographs on LCD screens," in *Proc. IEEE Int. Conf. Acoust. Speech Signal Process. (ICASSP)*, Mar. 2010, pp. 1790–1793.

[65] O. Bimber and D. Iwai, "Superimposing dynamic range," *ACM Trans. Graph.*, vol. 27, no. 5, 2008, Art. no. 150.

[66] S. Winkler, "Perceptual distortion metric for digital color video," *Proc. SPIE*, vol. 3644, pp. 175–184, May 1999.

[67] K. Bahrami and A. C. Kot, "A fast approach for no-reference image sharpness assessment based on maximum local variation," *IEEE Signal Process. Lett.*, vol. 21, no. 6, pp. 751–755, Jun. 2014.

[68] R. Nosaka, Y. Ohkawa, and K. Fukui, "Feature extraction based on co-occurrence of adjacent local binary patterns," in *Proc. Pacific-Rim Symp. Image Video Technol.*, 2011, pp. 82–91.

[69] V. Ojansivu and J. Heikkilä, "Blur insensitive texture classification using local phase quantization," in *Proc. Int. Conf. Image Signal Process.*, 2008, pp. 236–243.
[70] J. Kannala and E. Rahtu, "BSIF: Binarized statistical image features," in *Proc. IEEE Int. Conf. Pattern Recognit. (ICPR)*, Nov. 2012, pp. 1363–1366.
[71] D. G. Lowe, "Distinctive image features from scale-invariant keypoints," *Int. J. Comput. Vis.*, vol. 60, no. 2, pp. 91–110, 2004.
[72] J. Yosinski, J. Clune, Y. Bengio, and H. Lipson, "How transferable are features in deep neural networks?" in *Proc. Adv. Neural Inf. Process. Syst.*, 2014, pp. 3320–3328.
[73] I. Chingovska, A. Anjos, and S. Marcel, "On the effectiveness of local binary patterns in face anti-spoofing," in *Proc. IEEE Int. Conf. Biometrics Special Interest Group (BIOSIG)*, Sep. 2012, pp. 1–7.
[74] P. Viola and M. Jones, "Rapid object detection using a boosted cascade of simple features," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, vol. 1. Dec. 2001, pp. I-511–I-518.
[75] S. Ben-David, J. Blitzer, K. Crammer, A. Kulesza, F. Pereira, and J. W. Vaughan, "A theory of learning from different domains," *Mach. Learn.*, vol. 79, nos. 1–2, pp. 151–175, May 2010.

Haoliang Li received the B.S. degree from the University of Electronic Science and Technology of China in 2013. He is currently pursuing the Ph.D. degree with Nanyang Technological University, Singapore. His research interest include multimedia forensics.

Wen Li received the B.S. and M.Eng. degrees from the Beijing Normal University, Beijing, China, in 2007 and 2010, respectively, and the Ph.D. degree from Nanyang Technological University, Singapore, in 2015. He is a Post-Doctoral Researcher with the Computer Vision Laboratory, ETH Zürich, Switzerland. His main interests include transfer learning, multi-view learning, multiple kernel learning, and their applications in computer vision.

Hong Cao is the Head of data science for Ernst & Young Advisory Pte Ltd., (EY Advisory). He is responsible for setting up and managing EYs ASEAN Data Science Team, and leading the data science projects through the pipeline starting from presales, scoping to delivery. He is also a Data Science Evangelist and advises the higher management on the current data science technologies and the undertaken strategy. He has published about 50 scientific publications in top-tier venues of machine learning, data mining, and signal processing. His research in data-driven image forensics received the Best Paper Award in IWDW 2010 and an honorary mention in ISCAS 2010. He is also a winner of multiple international data science competitions, such as GE Flight Quest in 2013 and OPPORTUNITY activity recognition challenge in 2011. He currently chairs the IEEE Signal Processing Society, Singapore Chapter.

Shiqi Wang (M'15) received the B.S. degree in computer science from the Harbin Institute of Technology in 2008 and the Ph.D. degree in computer application technology from Peking University in 2014. From 2014 to 2016, he was a Post-Doctoral Fellow with the Department of Electrical and Computer Engineering, University of Waterloo, Waterloo, ON, Canada. From 2016 to 2017, he was with the Rapid-Rich Object Search Laboratory, Nanyang Technological University, Singapore, as a Research Fellow. He is currently an Assistant Professor with the Department of Computer Science, City University of Hong Kong. He has proposed over 30 technical proposals to ISO/MPEG, ITU-T, and AVS standards. His research interests include video compression, image/video quality assessment, and image/video search and analysis.

Feiyue Huang graduated from Tsinghua University in 2008 and received the Ph.D. degree in computer science. He is the Director and an Expert Researcher with the Tencent Youtu Laboratory. He joined Tencent and engaged in technical research, such as computer vision and machine learning in 2008. In 2012, he founded the Youtu Team and acquired many leading technology achievements in the fields of face recognition and image recognition.

Alex C. Kot (S'85–M'89–SM'98–F'06) has been with Nanyang Technological University, Singapore, since 1991. He headed the Division of Information Engineering, School of Electrical and Electronic Engineering for eight years and served as an Associate Chair/Research. He was the Vice Dean Research with the School of Electrical and Electronic Engineering and the Associate Dean for the College of Engineering for eight years. He is currently a Professor with the School of Electrical and Electronic Engineering and the Director of the Rapid-Rich Object Search Lab. He has published extensively in the areas of signal processing for communication, biometrics, image forensics, information security, and computer vision.

He is a fellow of the IES and the Academy of Engineering, Singapore. He received the IEEE Distinguished Lecturer of the Signal Processing Society. He was a recipient of the Best Teacher of the Year Award. He has co-authored several best paper awards, including for ICPR, IEEE WIFS, and IWDW. He has served the IEEE Signal Processing Society in various capacities, such as the General Co-Chair at the 2004 IEEE International Conference on Image Processing and the Vice President of the IEEE Signal Processing Society. He served as an Associate Editor for the IEEE TRANSACTIONS ON SIGNAL PROCESSING, the IEEE TRANSACTIONS ON MULTIMEDIA, the IEEE SIGNAL PROCESSING LETTERS, the *IEEE Signal Processing Magazine*, the IEEE JOURNAL OF SPECIAL TOPICS IN SIGNAL PROCESSING, the IEEE TRANSACTIONS ON CIRCUITS AND SYSTEMS FOR VIDEO TECHNOLOGY, the IEEE TRANSACTIONS ON INFORMATION FORENSICS AND SECURITY, the IEEE TRANSACTIONS ON IMAGE PROCESSING, the IEEE TRANSACTIONS ON CIRCUITS AND SYSTEMS I: FUNDAMENTAL THEORY AND APPLICATIONS, and the IEEE TRANSACTIONS ON CIRCUITS AND SYSTEMS II: ANALOG AND DIGITAL SIGNAL PROCESSING.