Open Set Fingerprint Spoof Detection Across Novel Fabrication Materials

Ajita Rattani, Walter J. Scheirer and Arun Ross

***Abstract*—A fingerprint spoof detector is a pattern classifier that is used to distinguish a live finger from a fake (spoof) one in the context of an automated fingerprint recognition system. Most spoof detectors are learning-based and rely on a set of training images. Consequently, the performance of any such spoof detector significantly degrades when encountering spoofs fabricated using novel materials not found in the training set. In real-world applications, the problem of fingerprint spoof detection must be treated as an open set recognition problem where incomplete knowledge of the fabrication materials used to generate spoofs is present at training time, and novel materials may be encountered during system deployment. To mitigate the security risk posed by novel spoofs, this work introduces: (a) the use of the Weibull-calibrated SVM (W-SVM), which is relatively robust for open set recognition, as a novel-material detector and a spoof detector, and (b) a scheme for the automatic adaptation of the W-SVM-based spoof detector to new spoof materials that leverages interoperability across classifiers. Experiments conducted on new partitions of the LivDet 2011 database designed for open set evaluation suggest (i) a 97% increase in the error rate of existing spoof detectors when tested using new spoof materials, and (ii) up to 44% improvement in spoof detection performance across spoof materials when the proposed adaptive approach is used.**

***Index Terms*—Fingerprint Spoofing, Spoof Detection, Presentation Attacks, Statistical Learning, Open Set Recognition.**

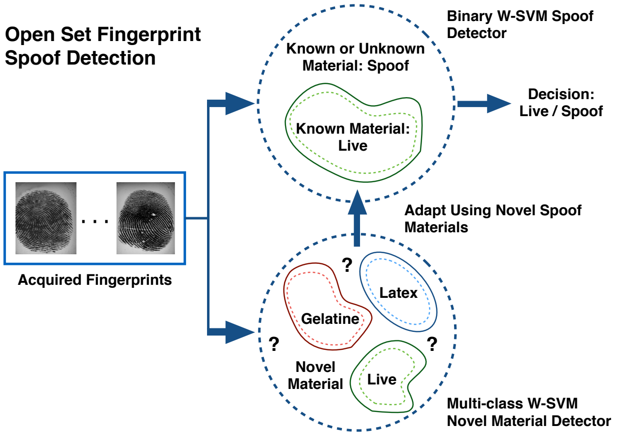
Open Set Fingerprint Spoof Detection
Binary W-SVM Spoof Detector
Known or Unknown Material: Spoof
Known Material: Live
Decision: Live / Spoof
Acquired Fingerprints
Adapt Using Novel Spoof Materials
Gelatine
Latex
Novel Material
Live
?
?
?
Multi-class W-SVM Novel Material Detector

Fig. 1: Illustration of the proposed scheme for automatic detection and adaptation of a fingerprint spoof detector to new spoof materials. Input samples detected as new spoof materials by the novel-material detector are used to adapt the spoof detector. Both the novel-material detector and spoof detector are implemented using the Weibull-calibrated SVM (W-SVM) [39], an algorithm designed for open set recognition.

I. Introduction

The history of fingerprint spoofing in the field of forensics is almost as old as that of fingerprint classification itself. In fact, the question of whether or not fingerprints left behind in a crime scene could be forged was answered in the affirmative in 1924 [47], before it was even formally posed as a question in 1936 [8]. A recurring theme in the historical record is the use of new spoofing materials and techniques to thwart methods specifically designed to prevent fingerprint spoofing.

In the field of biometrics, a spoofing attack occurs when an attacker mimics the biometric trait of another individual to circumvent a biometric authentication system. For instance, a fake finger can be fabricated using commonly available materials such as latex, glue, and gelatin, with the fingerprint ridges of an individual engraved on the surface [26], [48], [24], [2]. An attacker can place the fake finger on a fingerprint sensor and claim the identity of the owner of the actual ridges. Such attacks pose a direct threat because they leverage commonly available materials and do not require any knowledge of the internal functionality of the underlying biometric authentication system. The success rate of this kind of fingerprint spoof attack can be above 70% [26], [4].

But another danger is in emerging spoof attacks that are not as preventable as those that take advantage of known materials. Practical evidence of novel fingerprint spoofing attacks is mounting up. (1) In 2008, a South Korean woman was caught trying to pass through the immigration screening system in Nippon, Japan by using a special tape with someone else's fingerprints on her fingers to fool the fingerprint recognition machine[1]; (2) Similarly in 2013, a Brazilian doctor was arrested in São Paulo for using prosthetic silicone fingers to fool the biometric device that tracks employee attendance at the hospital where she worked[2]; and (3) Shortly after the release of Apple's iPhone5S in 2013, the German hacker group Chaos Computer Club spoofed its fingerprint scanner with a hybrid combination of materials[3].

It is often straightforward to create spoofs given a source fingerprint of even modest quality. Prints can be obtained via (a) the consensual method (*i.e.*, with the collaboration of the user) [48], (b) the non-consensual method [14] (a

A. Rattani is with the Dept. of CSE, University of Missouri-Kansas City, USA. email: rattania@umkc.edu.

W. J. Scheirer is with the Dept. of CSE, University of Notre Dame, USA. email: wscheire@nd.edu.

A. Ross is with the Dept. of CSE, Michigan State University, USA. Corresponding author: A. Ross (email: rossarun@cse.msu.edu).

[1] http://www.smh.com.au/travel/womanfools-japans-airport-security-fingerprint-system-20090102-78rv.html.

[2] http://www.bbc.com/news/world-latin-america-21756709

[3] http://www.ccc.de/en/updates/2013/ccc-breaks-apple-touchid

latent fingerprint is lifted using specialized tools), or (c) by reverse engineering the minutiae template from a biometric authentication system [38]. In all three cases, the fake fingerprint fabrication process typically consists of the following steps: (1) a mould is created from the source print; (2) *any* liquid casting (fabrication) material suitable for contact with a fingerprint sensor is poured on the mould; and (3) after the liquid solidifies, the cast is lifted from the mould and is used as a fingerprint replica or fake finger. The flexibility in material choice afforded by step 2 is good news for an attacker intent on evading detection, but a frustrating confound for designers of robust anti-spoofing algorithms.

Fingerprint spoof detection algorithms are the first line of defense against such attacks on fingerprint authentication systems [48], [21], [24]. Existing fingerprint spoof detection algorithms extract textural features (such as local binary patterns [29]), coarseness features (statistics from residual noise [27]), anatomical features (such as pore details [10]) or physiological attributes (such as perspiration [1]) from live and fake fingerprint samples to train a binary classifier (*e.g.*, a Support Vector Machine). The output of a spoof detection algorithm consists of a label: “Live” or “Spoof”. The output may also include a score reflecting the probability that a fingerprint sample corresponds to a real live finger.

To evaluate the effectiveness of existing fingerprint spoof detection algorithms, the biometrics research community has organized a regular series of competitions (LivDet) since 2009. Despite recent advances in machine learning-based approaches, the state-of-the-art in fingerprint spoof detection is not mature enough to be deployed in real-world systems. This is because existing fingerprint spoof detection algorithms do not exhibit acceptable error rates [48], [34], [15], [35], [24] – even when they are trained and tested on the same set of fabrication materials, *i.e.*, *closed set recognition*. In learning-based algorithms, performance is significantly influenced by the fabrication materials used to generate spoofs during the training stage. Reported studies [44], [48], [22] suggest a *three* fold increase in the error rates of fingerprint spoof detectors when spoofs using new materials (not used during the training stage) are encountered during the testing or operational stage, *i.e.*, *open set recognition*. This means the generalization capability of existing fingerprint spoof detectors is limited across materials.

As spoofing attacks evolve, new materials will be used to launch spoof attacks. Given that it is not possible to train the spoof detector with spoofs generated from all possible fabrication materials [37], the problem of fingerprint spoof detection must be treated as an open set recognition problem [42] where spoofs generated using novel materials that are not known during the training stage are encountered during the testing stage. *The aim of this work is to design a scheme for the automatic detection and adaptation of the spoof detector to novel-material spoofs*. Specifically, input fingerprint samples detected as new spoof materials (not known during the training stage) by the novel material detector, are used to adapt the fingerprint spoof detector. Both the novel-material and spoof detectors are implemented using Weibull-calibrated SVM (W-SVM) [42]. Fig. 1 shows the schema of the proposed adaptation scheme implemented using W-SVM. Such a scheme should (a) significantly reduce error rates on spoof samples in the test set that are generated using new materials, and (b) preempt the need to perform supervised re-training of the spoof detector to cope with the advancement of spoofing techniques using novel materials.

A preliminary version of this article appeared at the International Joint Conference on Biometrics (IJCB) in 2014 [36]. In [36], a novel material detector was designed using an AdaBoost-based classifier that automatically detects and adapts an SVM-based spoof detector to new spoof materials. The contributions of this work over [36] are as follows:

- Unlike previous work, explicitly posing the problem as an open set recognition problem, thereby lending itself to a rigorous framework based on open set classifiers.
- The first application of the Weibull-calibrated SVM (W-SVM) to the problem of novel material detection and fingerprint spoof detection. The W-SVM makes use of recent advances in extreme value theory statistics for machine learning to directly address the risk of the unknown in an open set recognition problem.
- The automatic adaptation of the W-SVM-based fingerprint spoof detectors to new spoof materials. In contrast to [36], where AdaBoost-based classifiers were used for novel-material detection and SVM-based classifiers were used for spoof detection, the W-SVM can be used for both tasks and supports interoperability between individual detectors.
- An exhaustive experimental analysis incorporating all four sensors (Biometrika, Italdata, DigitalPersona and Sagem) in the LivDet 2011 dataset. Further, a baseline comparative analysis with the performance of the spoof-detector trained with the ground-truth (oracle test) is also provided.

II. Prior Work on Open Set Spoof Detection

Tan et al. [44] evaluated the impact of novel spoof materials on the performance of fingerprint spoof detection. A fingerprint spoof detector based on ridge signal and valley noise features was trained using play-doh, gelatin, and silicone, and tested using latex rubber, latex caulk, and latex paint. An equal error rate (EER) of 3.5%, 5.9% and 5.8%, respectively, was achieved when testing on new instances of the training materials for Identix, Crossmatch, and DigitalPersona sensors. Results showed an increase in error to 14.5%, 55.6% and 36.6%, respectively, when novel spoof materials were used during the testing stage.

Marasco and Sansone [22] also evaluated the impact of novel spoof materials on fingerprint spoof detection algorithms based on coarseness, texture, perspiration, morphology and various combinations of these features. Algorithms were trained on samples generated using either gelatin, play-doh or silicone from the LivDet 2009 database (Identix and Crossmatch sensors), and tested on the other two materials. The accuracy of all the algorithms dropped across spoof fabrication materials by about 24%. However, the algorithm which fused perspiration and morphology-based features outperformed the others.

The two studies above evaluated the impact of novel spoof fabrication materials on spoof detection, but they did not propose any solution to the problem. In a recent study, Rattani and Ross [37] devised a scheme to improve the interoperability of spoof detectors across spoof fabrication materials. In their study, a pre-processing scheme is proposed based on linear and non-linear denoising in order to reduce the differences in noise levels (surface coarseness) in fake fingerprint images corresponding to different types of fabrication materials. In this regard, a combination of linear filtering (gaussian filter) as well as non-linear image denoising (symlet-based wavelet) is employed before the fingerprint spoof detector is invoked. The proposed denoising scheme was observed to enhance the generalization ability of an LBP-based spoof detector across spoof materials by up to 44% on the LivDet 2011 database [48]

Further, Rattani and Ross [36] proposed a scheme for the automatic detection and adaptation of the spoof detector to spoofs fabricated using novel materials that are encountered after system deployment. To this end, a novel-material detector, implemented using AdaBoost, was developed that detects spoofs made of new materials. Samples flagged as new spoofs were used to *automatically* retrain and update an SVM-based fingerprint spoof detector. The proposed automatic novel-material detection scheme obtained an average correct detection rate of up to 74% on a partition of the LivDet 2011 database (corresponding to the Biometrika sensor) designed for open set evaluation. The performance of the fingerprint spoof detector when retrained based on the output of the novel material detector was observed to improve by up to 46%. However, the proposed method adds to the computational overhead of the overall system. This is due to the use of two different classifiers: multi-class Adaboost for novel material detection and binary SVM for spoof detection.

III. Features used for Spoof Detection

A fingerprint spoof detector aims to disambiguate real live fingerprints from fake fingerprints by exploiting their differences in textural, physiological and anatomical attributes. Features based on these attributes are extracted from the training set of live and fake fingerprint samples, and a binary classifier (such as SVM) is learned. The output of the spoof detection algorithm is often a numerical value, called a liveness measure, indicating the probability that the input fingerprint sample corresponds to a live finger. Table I shows fingerprint attributes and the associated features proposed in the literature for the task of fingerprint spoof detection.

In comparative evaluations on the LivDet 2011 database [10], [12], local textural features (such as LBP, LPQ and BSIF) have been shown to outperform other competing spoof measures based on anatomical features (such as pores [25]) and perspiration [1], as well as other algorithms anonymously submitted to that challenge whose error rates were in the range [20%, 40%]. This suggests *the efficacy of local textural descriptors* in detecting the difference in the texture between live and fake fingerprints, which is caused by loss of information and errors introduced during the fake fingerprint fabrication process.

TABLE I: Examples of attributes and the associated features used in existing studies on fingerprint spoof detection.

Attributes	**Associated studies and features used in fingerprint spoof detection**
Coarseness	Moon et al.'s coarseness analysis using noise residue [27] Coli et al.'s power spectrum analysis [7] Tan and Schukers' wavelet-based statistics [45]
Perspiration	Abhyankar and Schukers' perspiration analysis using wavelets [1] Marasco and Sansone's fusion of morph. and perspiration analysis [23]
Anatomical	Marcialis et al.'s statistics related to fingerprint pore analysis [25] Espinoza and Champod's pore analysis for spoof detection [9] Tan and Schukers' fusion of ridge signal and valley noise analysis [46]
Textural	Nikam and Agarwal's grey level co-occurence matrix (GLCM) [30] Nikam and Agarwal's local binary patterns (LBP) [29] Ghiani et al.'s local phase quantization (LPQ) [13] Jia et al.'s local ternary patterns (LTP) [18] Sansone at al.'s weber local descriptors (WLD) [15] Ghiani et al.'s binary statistical image features (BSIF) [11]

IV. Fabrication Materials Used for Spoofing

A variety of readily available materials such as latex, gelatin, silicone, play-doh, etc., have been used to fabricate fake fingerprints and circumvent fingerprint sensors operating based on optical, capacitive and other principles [19], [48]. Optical sensors are susceptible to spoof attacks when the fabrication material used has a light reflectivity similar to that of skin. Capacitive scanners can be fooled by the use of inherently conductive spoof materials such as gelatin, glycerin or wood glue[4].

The casting (*i.e.*, fabrication) material should have high elasticity and very low shrinkage to avoid reduction in volume as the cast cools and solidifies. In fact, more than *fifty seven materials and material variants* have been identified for fake fingerprint fabrication [31]. However, different materials exhibit different characteristics:

- *Differences in artifacts*: Different fabrication materials possess different potentials to hold a ridge and valley pattern. This can result in fabrication errors. Further, due to differences in the elasticity of the materials, non-linear deformations may be introduced when pressure is applied while presenting the fake finger to the sensor. Fig. 2 shows example of fake fingerprint samples corresponding to five different fabrication materials (from LivDet 2011 database [48]). Fabrication errors and non-linear deformations (examples indicated by the red circle and white square) are quite evident in the case of silgum, wood glue and ecoflex.
- *Differences in contrast between ridges and valleys*: Fig. 2 also shows the difference in contrast between the fake fingerprints fabricated using five different materials for a subject in the LivDet 2011 database. Low contrast is evident for silgum and wood glue, while high contrast is evident for latex and gelatin.
- *Differences in surface coarseness*: Due to the presence of organic molecules in fabrication materials that tend to agglomerate, noise components are observed in fake fingerprint images [27]. As a consequence, the surface of a fake fingerprint is coarser than its live counterpart. Further, the coarseness varies across different fabrication materials [37].

[4] http://nexidbiometrics.com/faq/

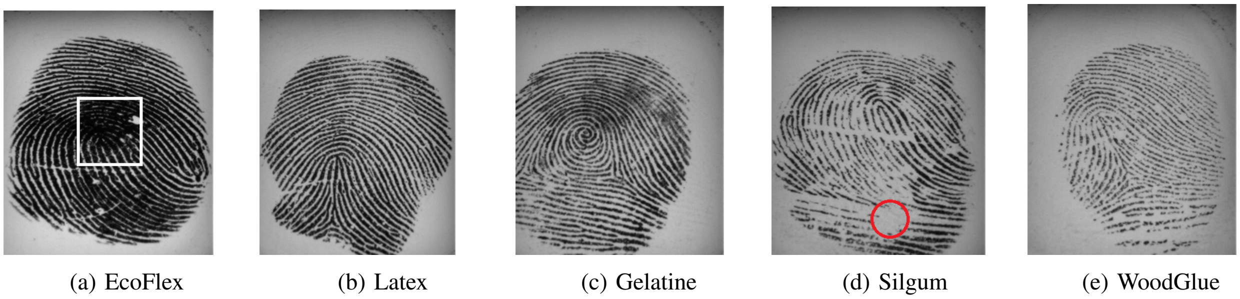

Fig. 2: Examples of fake fingerprint images (from the LivDet 2011 [48] database) corresponding to five different fabrication materials. The artifacts introduced (examples indicated by the circle and square) are typically quite prominent for silgum, wood glue and ecoflex materials.

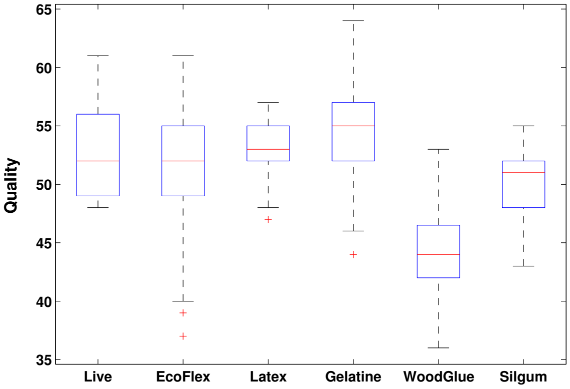

Fig. 3: Quality measures computed for 200 live and 200 fake fingerprint samples (acquired using Biometrika sensor) fabricated using five different materials from LivDet 2011 [48].

- *Differences in image quality*: As a consequence of the above factors, the quality of the fake fingerprint samples may vary across fabrication materials. Fig. 3 shows the difference in the range of quality values (computed using the Image Quality of Fingerprint (IQF) software from MITRE[5]) across spoof samples generated using different fabrication materials. It can be seen that silgum and wood glue produced spoofs of relatively low quality. In contrast, the quality of latex spoofs is quite similar to that of live fingerprint samples.

Due to the aforementioned reasons, the performance of a spoof detection algorithm degrades when spoofs generated using new materials are encountered during operation. This highlights the need for dynamically adapting the spoof detector to new spoof materials during the operational phase. Next, we explain the proposed scheme for automatic detection and adaptation of the spoof detector to new spoof materials, supporting open set spoof detection.

V. Interoperable Novel Material Detection and Spoof Detection Based on the W-SVM

Prior work on spoof detection in an open set context [36] broke the problem up into two distinct tasks, each addressed by a different supervised learning algorithm. The task of novel-material detection was approached via AdaBoost, while SVM proved to be effective for spoof detection. These algorithms were chosen for their empirical performance, rather than a specific theoretical property related to open set recognition. In this article, we suggest that a unified approach is more attractive for this problem because it operates on the same input feature space for both tasks. Moreover, by making use of a learning approach grounded in a strong theory that directly addresses open set recognition, we can generalize beyond the baseline performance achieved by AdaBoost and SVM. Thus, we turn to emerging work in machine learning on statistical extreme value theory and its relationship to the open set recognition problem.

A naïve way to solve an open set problem like novel-material detection or spoof detection is to simply set an empirically estimated threshold over a distance, calculated by a Nearest Neighbor (NN) algorithm, between an input print and the closest matching print in a database. However, such thresholds estimated over training or validation data do not generalize well, since sufficiently dense samples in training are not always available, and for multi-class open set problems like novel-material detection, the distance space is inconsistent across classes. AdaBoost and SVM are typically more powerful that NN approaches, in that they learn decision models over many labeled training points using more sophisticated statistical strategies (combinations of weak learners in the case of AdaBoost, and maximum margin in the case of SVM). But this again returns to the problem of estimating a threshold over inconsistent score spaces in order to use these algorithms.

Calibration is one possible solution to this problem. By enforcing consistency between all distances or scores across all spoof material classes and the live class, better decision models can be deployed (*e.g.*, a threshold over interpretable probabilities). Most commonly, models that fit all of the scores derived from the training data are used for this purpose [33], [49], [28]. However, algorithms like SVM produce models that are composed of a subset of the training data (the support vectors), which leads to a strongly discriminative representation for recognition that is just the edge of the class distribution. As an alternative to fitting all of the training data, distributions from the statistical extreme value theory (EVT) [6] family can be used to only fit data near the decision boundary. Theoretical

[5] http://www2.mitre.org/tech/mtf/

work on recognition algorithms has shown that the recognition problem itself is consistent with the assumptions of EVT [41], yielding a useful tool to generate probabilities, regardless of the overall distribution of data.

Scheirer et al. have proposed several techniques for EVT-based SVM calibration [40], [17], [39]. Out of these, the best performing algorithm for open set recognition problems is currently the W-SVM [39], a Weibull-calibrated formulation that combines a 1-Class SVM with a binary SVM, both with non-linear kernels. Why does such an algorithm help for open set problems like novel-material detection and spoof detection? First, when Weibull modeling is coupled with a 1-Class SVM using a radial basis function kernel, it can be proved that the probability of class membership decreases in value as points move from known data toward open space. Second, the Weibull distribution provides better modeling at the decision boundaries for a binary SVM, resulting in good generalization even in the presence of many unknown classes. Novel material detection and spoof detection are difficult problems because there are often small inter-class distances in the feature space – effective spoofs are designed to mimic live skin. The W-SVM ensures that the probability models do not treat data at the decision boundaries as low probability members of a class, where separation between spoof material classes and the live class in a raw distance sense may be limited. In the rest of this section, we will describe the operation of the W-SVM algorithm specifically for novel-material detection and spoof detection.

A. Training a W-SVM for Novel-Material Detection and Spoof Detection

The W-SVM training algorithm consists of four distinct steps split into two different classification regimes: 1-Class and Binary. The base formulation applies to multi-class classification problems such as novel material detection, and binary classification problems such as spoof detection. Source code for the implementation described below is publicly available on GitHub[6].

1) 1-Class RBF SVM Training: The first step of W-SVM training is to train a 1-Class SVM [43]. With the absence of a second class in the training data, the origin defined by a kernel function Ψ serves as the only member of a "second class." The objective of the 1-Class SVM is to find the best margin with respect to the origin. The resulting binary classification function f^o after training takes the value $+1$ in a region capturing most of the training data points, and -1 elsewhere. For a multi-class problem like novel-material detection, an individual classifier can be trained for each known labeled spoof material and the live skin class, yielding a set of classifiers $f_1^o, \ldots, f_n^o$.

Let $p(x)$ be the probability density function estimated from the training data $\{x_1, x_2, \ldots, x_m \mid x_i \in X\}$, where X is a single class. A mapping function $\Phi : X \to H$ transforms the training data into a different space. To separate the training data from the origin, the algorithm solves the following quadratic programming problem for w and ρ to learn f:

$$min \frac{1}{2} \| w \|^2 + \frac{1}{\nu m} \sum_{i=1}^{l} \xi_i - \rho \tag{1}$$

subject to

$$(w \cdot \Phi(x_i)) \geq \rho - \xi_i \quad i = 1, 2, \ldots, m \quad \xi_i \geq 0 \tag{2}$$

where ρ is an offset that parameterizes the hyperplane in the feature space defined by the mapping Φ, and ξ_i are slack variables. The inner product in the image of Φ can be computed by evaluating a kernel Ψ.

$$\Psi(\mathbf{x}, \mathbf{x}') = (\Phi(\mathbf{x}) \cdot \Phi(\mathbf{x}')) \tag{3}$$

For the W-SVM, Ψ is a radial basis function, which impacts density estimation and smoothness in a parameterized fashion:

$$\Psi(\mathbf{x}, \mathbf{x}') = \exp(-\gamma ||\mathbf{x} - \mathbf{x}'||^2), \gamma > 0 \tag{4}$$

The regularization parameter $\nu \in (0, 1]$ controls the trade-off between training classification accuracy and the smoothness term $\| w \|$, and also impacts the choice and number of support vectors. In the 1-Class SVM, $p(x)$ is cut by the margin plane minimizing Eq. 1 and satisfying Eq. 2. Regions of $p(x)$ above the margin plane define positive classification and capture most of the training data.

In effect, the margin plane serves as a *de facto* threshold, which partially addresses the problem of threshold estimation described above. The effectiveness of f^o is dependent on how the 1-Class parameter ν and RBF parameter γ are set. In plain language, ν is an upper bound on training error, while γ controls the extent of influence for a single training example. The higher the value is for ν, the more tolerance there is for misclassification during training. The lower the value is for γ, the larger the positive decision region will be. Importantly, a non-linear kernel like RBF limits our open space risk (*i.e.*, the risk of the unknown) by eliminating the half-space problem for linear classifiers [39], where space far from the support of known positive training samples is always assigned a positive class label. This is illustrated in Figs. 4 & 5, where it can be seen that the positive decision space is finite.

2) 1-Class RBF SVM EVT Calibration: Turning to calibration in the second step of W-SVM training, the probability of class inclusion for a 1-Class SVM can be modeled by fitting a Weibull distribution to scores generated by classifying the training data $\{x_1, x_2, \ldots, x_m\}$ using the corresponding trained model f^o. This provides a set of scores S However, the extrema from the overall score distribution are of interest for modeling. Thus, let $O \subset S$ be the lower tail of the scores, not exceeding 50% of the overall number of scores; [39] recommends choosing a tail size of 0.5× or 1.5× the number of support vectors defining f^o, depending on the problem.

A Weibull is the expected distribution for the lower tail because it is bounded from below. A Weibull distribution has three parameters: location ς, scale λ, and shape κ. Maximum Likelihood Estimation (MLE) can be applied to estimate the $\varsigma_o, \lambda_o, \kappa_o$ that best fit O. To calculate the probability of class

[6] https://github.com/ljain2/libsvm-openset

inclusion for a particular SVM decision $f^o(x)$, the CDF defined by the parameters is used:

$$P_O(y|f^o(x)) = 1 - e^{-(\frac{f^o(x)-\varsigma_o}{\lambda_o})^{\kappa_o}}. \tag{5}$$

This calibration model serves as a conditioner: if the 1-Class SVM predicts $P_O(y|f^o(x)) > \delta_\tau$, even with a very low threshold δ_τ, that a given input x is a member of class y, then we will consider the binary classifier's estimates. A rejection at this step is the first way we can detect a novel spoof.

3) Binary RBF SVM Training: The 1-Class SVM reduces the open set risk of any problem to 0 (see the proofs in Sec. 3 of [39]). However, it is well known that the 1-Class formulation tends to overfit the training data from the positive class [20], [51], [42]. Some knowledge of known negative classes during training improves discrimination by enforcing separation between known classes. For the application of novel material detection, this means that when a spoof detector is trained with positive samples from the live class, and negative samples from multiple spoof material classes, it can generalize to unseen samples from the known classes much more effectively than the 1-Class SVM.

Binary SVMs attempt to learn a margin that maximizes the separation between two classes. Let $\boldsymbol{\alpha}$ be a set of Lagrange multipliers. To separate the data in the non-linear case, the following (dual) optimization problem for the C-SVM [5] formulation can be solved:

$$\begin{aligned} \max_{\alpha_i \geq 0} \quad & \sum_i \alpha_i - \frac{1}{2}\sum_{j,k} \alpha_j \alpha_k y_j y_k \Psi(x_j, x_k) \\ \text{subject to} \quad & 0 \leq \alpha_i \leq C, \forall i; \\ & \sum_i \alpha_i y_i \end{aligned} \tag{6}$$

where x_i is the i-th training example from the data $\{x_1, x_2, \ldots, x_m \mid x_i \in X\}$, X contains positive and negative samples, C is the soft margin parameter, and $y_i \in \{-1, +1\}$ is, for the i-th training example, the correct output label. As in the 1-Class case discussed above, Ψ is the RBF kernel defined in Eq. 4. The resulting binary SVM function is denoted as f.

4) Binary RBF SVM EVT Calibration: We also require calibration at the decision boundary in the binary case. However, different from the 1-Class case, EVT distributions are fit separately to the positive and the negative scores from the binary SVM. The positive class modeling proceeds as it did above with a Weibull distribution, but a reverse Weibull is used for the largest scores from the negative examples because they are bounded from above. In the context of novel material detection or liveness detection, binary SVM calibration can be formulated as follows. For the purpose of demonstration, assume the training examples are separated into positive examples of a live skin class, $x \in \mathcal{K}^+$, and negative examples from all other known spoof materials, $x \in \mathcal{K}^-$. Letting $s_i = f(x_i)$ be the SVM decision score for x_i, scores are collected into live skin and spoof material sets where scores for live skin are $s_j \in S^+$ if $x_j \in \mathcal{K}^+$ and scores for spoof materials are $s_j \in S^-$ if $x_j \in \mathcal{K}^-$. Let ψ be the upper extremes of the scores S^- from the spoof material classes, and let η be the lower extremes of the scores S^+ from the live skin class.

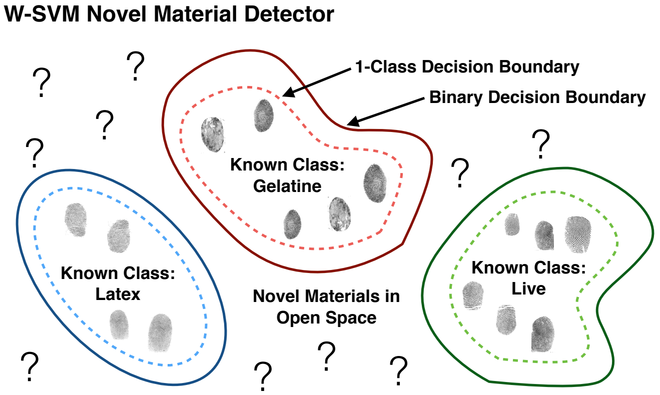
W-SVM Novel Material Detector
1-Class Decision Boundary
Binary Decision Boundary
Known Class: Gelatine
Known Class: Latex
Known Class: Live
Novel Materials in Open Space
?

Fig. 4: The W-SVM algorithm is by design a multi-class supervised learning method [39], making it amenable to the task of novel-material detection. The objective of this task is to determine whether or not a fingerprint image belongs to a known class (*e.g.*, Live, Gelatine or Latex), or is a novel point in open space (*i.e.*, belongs to class "other"). The detection of novel spoofing materials is useful for adapting liveness detectors to prevent new attacks. Notice that the algorithm establishes two decision boundaries for each known class model: a 1-Class decision boundary to reduce open space risk, and a binary decision boundary to improve classification accuracy via generalization. Both decision models are calibrated via the statistical extreme value theory.

MLE can be used to estimate the $\varsigma_\eta, \lambda_\eta, \kappa_\eta$ that best fit η and the $\varsigma_\psi, \lambda_\psi, \kappa_\psi$ that best fit ψ. To produce a probability score for a particular SVM decision $f(x)$, the CDF defined by the parameters is used. Given a test sample x, two independent estimates for $P(y|f(x))$ are possible: P_η based on the Weibull CDF derived from the live skin class scores:

$$P_\eta(y|f(x)) = 1 - e^{-(\frac{f(x)-\varsigma_\eta}{\lambda_\eta})^{\kappa_\eta}} \tag{7}$$

and P_ψ based on the reverse Weibull CDF derived from the spoof material scores, which is equivalent to rejecting the Weibull fitting on the spoof material scores:

$$P_\psi(y|f(x)) = e^{-(\frac{f(x)-\varsigma_\psi}{\lambda_\psi})^{\kappa_\psi}}. \tag{8}$$

This same process can be applied for any combination of known live and spoof material classes. While P_η is not formally related to any 1-Class estimation, its use of only positive data means it shares some of the characteristics of 1-Class SVMs [17], including the ability to detect a novel spoof material. However, since the underlying classifier f is a one-vs-all binary SVM, the resulting estimates are more discriminative.

B. Novel-Material Detection with the W-SVM

To support novel-material detection, a multi-class W-SVM (Fig. 4) that is an ensemble of pairs of 1-Class and binary SVMs must be trained, where each pair detects a specific live skin or known spoof material class. A thresholding strategy that will allow for the rejection of unknown spoof materials is also required. Letting $P_O(y|x)$ be the probability from Eq. 5 for the RBF 1-Class SVM trained on positive examples of class y, an indicator variable can be defined as follows: $\iota_y = 1$ if $P_O(y|x) > \delta_\tau$ and $\iota_y = 0$ otherwise. Novel material detection for all known classes $\mathcal{Y}$ is then:

$$\begin{aligned} y^* = &\operatorname*{argmax}_{y\in\mathcal{Y}} P_{\eta,y}(x) \times P_{\psi,y}(x) \times \iota_y \\ &\text{subject to } P_{\eta,y^*}(x) \times P_{\psi,y^*}(x) \geq \delta_R. \end{aligned} \tag{9}$$

Notice that the novel-material detector has two parameters: δ_τ, which is generally very small (fixed to 0.001 for all experiments in this article) and is used to adjust what data the 1-Class SVM considers to be even remotely related to the positive class, and δ_R, which is the level of confidence needed in the W-SVM estimate itself. If both of these probabilities are not exceeded, a fingerprint sample is considered to be novel spoof. For the experiments in Sec. VI, The threshold δ_R is varied across the entire probability range to find the Equal Error Rate (EER) points for each classifier. The 1-Class parameter ν and RBF parameter γ for the 1-Class and binary SVMs are tuned to facilitate EER calculation, *i.e.*, values resulting in a smooth score space that contains an EER.

C. Spoof Detection with the W-SVM

To support spoof detection, we need to train a binary W-SVM (Fig. 5) that is able to discriminate between live skin and known spoof material classes. In this scenario, we again require a thresholding strategy that will allow us to reject unknown spoof materials. Assume that the live skin class is assigned the label "$+1$", and that any spoof class, known or unknown, is assigned the label "-1". Letting $P_O(+1|x)$ be the probability from Eq. 5 for the RBF 1-Class SVM trained on positive examples of the live skin class, we define an indicator variable: $\iota_{+1} = 1$ if $P_O(+1|x) > \delta_\tau$ and $\iota_{+1} = 0$ otherwise. spoof detection is then:

$$y = +1 \iff P_{\eta,+1}(x) \times P_{\psi,+1}(x) \times \iota_{+1} \geq \delta_R. \tag{10}$$

If the above condition isn't satisfied, $y = -1$. Similar to the novel material detector, δ_τ is set to a very low probability (0.001 for all experiments in this article) for spoof detection, and δ_R is again varied across the entire probability range to find the EER points for each classifier. The ν and γ parameters are tuned to produce a smooth score space.

D. Retraining a Spoof Detector Based on the Results of a Novel Material Detector

In a combined mode of operation, a spoof detector can be retrained based on the results of a novel material detector, ideally adapting to reject newly identified spoof materials. Since both algorithms are W-SVMs operating in the same feature space, the procedure is straightforward. For a set of feature vectors T from a collection of fingerprint images, feature vectors deemed by the novel material detector to not be members of any known classes (*i.e.*, fingerprint images receiving a probability score $\leq \delta_R$) are identified. These vectors are then added to the original set of training images as known negatives for the spoof detector, which is retrained using the procedure in Sec. V-A.

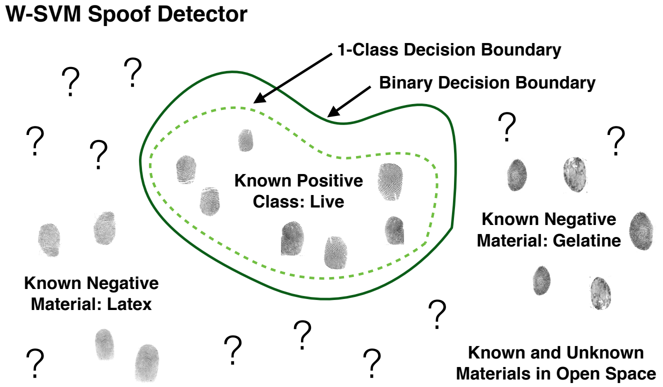
W-SVM Spoof Detector
1-Class Decision Boundary
Binary Decision Boundary
Known Positive Class: Live
Known Negative Material: Gelatine
Known Negative Material: Latex
Known and Unknown Materials in Open Space

Fig. 5: The W-SVM is also effective for the task of spoof detection, which unifies both components of our proposed spoof detection system. In this mode of operation, a binary classifier is trained with live prints as the positive class, and known spoofs as members of a single negative class. The classifier is able to reject prints created by both known and unknown spoof materials, which exist in the open space outside of the feature space of the live prints.

VI. Experimental Analysis

For all experiments, the LivDet 2011 [48] competition database was used. This database consist of $2,000$ live and $2,000$ spoof images from different subjects for four different acquisition sensors: Biometrika, Italdata, Sagem and DigitalPersona. The spoof images in the LivDet 2011 database were fabricated using the consensual method. Seven different spoof materials are available in the database: Latex, EcoFlex, Wood Glue, Gelatin, Silgum, Silicone and Play-Doh. However, not every sensor has corresponding images for the complete set.

Two other editions of LivDet are available (2009 and 2013), but LivDet 2011 is of particular interest for this experimental analysis because:

- It consists of a larger number of materials and sensors compared to LivDet 2009. Further, the quality of spoofs is higher in LivDet 2011, thus making open set spoof detection a more challenging task.
- As the spoofs are fabricated using the consensual method, it allows for the analysis of a difficult scenario in which high quality spoofs are fabricated via the help of a legitimate user and distributed among multiple users for illegitimate access. Spoofs in LivDet 2013 are fabricated using a non-consensual method for the Biometrika and Italdata sensors.
- Importantly, a number of spoof detection algorithms and comparative evaluations [10], [12], [13], [11] have been reported on LivDet 2011 in comparison to LivDet 2009 and the recently introduced LivDet 2013. In these evaluations and studies, error rates of the proposed spoof

detection algorithms are compared against the existing algorithms and those submitted (anonymously) to the LivDet 2011 competition. Thus, we have chosen the set which has attracted the most attention within the research community.

Protocol: The protocol used for the implementation and assessment of the proposed adaptive spoof detection scheme incorporating the W-SVM procedures of Sec. V is as follows. Note that the original protocol for LivDet 2011 has been modified to facilitate open set evaluation.

1) *Training*: For known class training data, the LivDet 2011 database is partitioned into $1,000$ live and 400 spoof images corresponding to **two** fabrication materials. Such sets are created for all combinations of materials for a sensor. Each binary classifier in a multi-class novel material detector $\mathcal{M}$ and each spoof detector $\mathcal{L}$ is trained from this data.
2) *Adaptation and Testing*: The test set of the LivDet 2011 database is divided into two non-overlapping partitions: T_1 and T_2. Each T_i consists of 500 live and 500 spoof samples. For the spoofs, 200 samples are from materials used during the training stage (known materials)[7] and 300 samples are from materials that were *not* used during the training stage (novel materials). First, the performance of a spoof detector trained on the training set is evaluated on T_1 and T_2. Next, the spoof detector is re-trained on those images in T_1 that are deemed to be a "novel spoof material" by an automatic novel material detector. Finally, the re-trained spoof detector is tested on T_2. In order to facilitate cross-validation, the roles of T_1 and T_2 are interchanged and the performance recomputed.

Performance assessment metrics:

1) *Novel material detector*: The performance of the novel material detector $\mathcal{M}$ is assessed using the following two metrics: (a) Correct detection rate (CDR), which is the proportion of spoof samples from novel materials that are correctly classified as novel materials, and (b) False detection rate (FDR), which is the proportion of live as well as spoof samples generated using known materials that are incorrectly classified as novel materials. The overall performance is summarized using the Equal Error Rate (EER$_{\mathcal{M}}$), which is the rate at which FDR = 1 - CDR.
2) *Fingerprint spoof detector*: The SVM output score from a spoof detector $\mathcal{L}$ is compared against a threshold to ascertain whether the input fingerprint is "Live" or "Spoof". The performance of the spoof detector is assessed again using EER. Here, EER is the rate at which the proportion of live samples incorrectly classified as fake is equal to the proportion of fake samples incorrectly classified as live.

Experiment #1: Performance of the spoof detector on previously known and novel materials: The goal of this experiment is to quantify the degradation in the performance of a W-SVM-based spoof detector when tested on spoof samples generated using novel materials. First, three different sets of spoof detectors are trained based on LBP, LPQ and BSIF features, respectively, extracted from 400 fake fingerprints corresponding to two materials and $1,000$ live fingerprint samples. The resulting sets of spoof detectors are denoted as $\mathcal{L}^{LBP}$, $\mathcal{L}^{LPQ}$ and $\mathcal{L}^{BSIF}$, respectively. The trained classifiers are tested on T_1 and T_2 consisting of a combined 400 spoof samples from known materials and 600 spoof samples from novel materials. Table II shows the EER of each spoof detector when tested using spoof samples from previously known (EER$_{known}$) and novel (EER$_{novel}$) materials.

The following observations can be made from Table II: (a) Different combinations of training materials exhibit different generalization performance. Often the combination of latex with other materials resulted in reduced error on the known as well as novel materials. (b) The average *increase* in the error rate of the spoof detector when tested on the set of novel materials is 93.6%, 78.1% and 237.9%, for BSIF, LBP and LPQ, respectively, averaged over four sensors: Biometrika, Italdata, DigitalPersona and Sagem. Training materials that tend to produce low quality spoofs, such as silgum and wood glue, lead to higher error rates when new spoof materials are encountered. (c) The average *increase* in the error rate when all three spoof detectors (LBP, LPQ and BSIF) are tested on new spoof materials is 51.9%, 49.0%, 156.2% and 182.9% on Biometrika, Italdata, DigitalPersona and Sagem sensors, respectively. This clearly conveys the need for designing a spoof detection scheme that is robust across fabrication materials.

Further, in comparison to results reported in [36] for the Biometrika sensor for nine material combinations, the average increase in error of the LBP- and LPQ-based spoof detectors is lower by 3.8% and 14.2%, respectively. For the BSIF-based spoof detector, average increase in the error on new spoof materials increased by 11.0%.

Experiment #2: Detection of new spoofs using a novel-material detector: This experiment evaluates the performance of the W-SVM-based novel material detection approach. In Table III, performance of the various textural descriptors used in the W-SVM framework (see Sec. V) for implementing the novel-material detector $\mathcal{M}$ is listed. These descriptors were extracted from live and spoof samples corresponding to two training materials; hence, $|\mathcal{Y}|$ = 3 classes for each multi-class detector $\mathcal{M}$. Novel material detection is implemented via the procedure described in Sec. V-B.

Table III reports EER, which is averaged over ten combinations of two training materials each. It can be seen that all the texture descriptors resulted in a high error rate, which can be attributed to the challenging nature of the open set recognition problem. LBP and BGP performed better than the other descriptors. The W-SVM-based novel material detector faired equally with the AdaBoost-based novel material detector proposed in [36]. The combination of LBP+BGP (two W-SVM classifiers were trained independently for the LBP and BGP descriptors, and their output was combined using the sum rule)

[7] Samples and subjects used in the training and test sets of LivDet 2011 do not overlap. Further, samples are not switched between training and test sets during performance evaluation in order to maintain the repeatability of the experiments.

TABLE II: Performance of the **BSIF**-, **LBP**- and **LPQ**-based spoof detectors when tested on previously known materials (EER$_{known}$) and on novel materials (EER$_{novel}$). The overall average increase in the error rate when encountering novel materials is 97.3%. This motivates the need for an adaptive algorithm. Full DET curves for the Biometrika + LBP combination can be found in the supplemental material.

Biometrika								
	$\mathcal{L}^{BSIF}$		$\mathcal{L}^{LBP}$		$\mathcal{L}^{LPQ}$		**Average**	
Training materials	**EER**$_{known}$ [%]	**EER**$_{novel}$ [%]	**EER**$_{known}$ [%]	**EER**$_{novel}$ [%]	**EER**$_{known}$ [%]	**EER**$_{novel}$ [%]	**EER**$_{known}$ [%]	**EER**$_{novel}$ [%]
Skin+Latex+EcoFlex	6.0	16.3	6.5	13.2	9.8	18.4	7.4	16.0
Skin+WoodGlue+Latex	15.0	15.0	10.0	13.8	14.4	16.8	13.1	15.2
Skin+Gelatine+Latex	11.0	16.5	12.0	11.2	8.9	17.7	10.6	15.1
Skin+Silgum+Latex	10.5	20.8	12.3	19.7	10.8	16.3	11.2	18.9
Skin+EcoFlex+Silgum	14.0	29.5	9.3	30.2	12.3	23.0	11.9	27.6
Skin+Gelatine+EcoFlex	13.3	23.3	9.7	15.2	14.0	22.4	12.3	20.3
Skin+Silgum+Gelatine	13.3	23.8	11.5	23.3	14.8	19.5	13.2	22.2
Skin+WoodGlue+Silgum	18.3	23.0	18.0	32.3	13.5	19.0	16.6	24.8
Skin+Gelatine+WoodGlue	16.8	17.2	12.3	11.0	15.8	17.3	15.0	15.2
Skin+WoodGlue+EcoFlex	16.3	17.2	21.7	26.7	17.4	17.3	18.5	20.4
Average EER ± STDERROR:	13.5 ± 1.1	**20.3** ± 1.5	12.3 ± 1.4	**19.7** ± 2.5	13.2 ± 0.9	**18.8** ± 0.7	12.9 ± 1.0	**19.6** ± 1.4

Italdata								
	$\mathcal{L}^{BSIF}$		$\mathcal{L}^{LBP}$		$\mathcal{L}^{LPQ}$		**Average**	
Training materials	**EER**$_{known}$ [%]	**EER**$_{novel}$ [%]	**EER**$_{known}$ [%]	**EER**$_{novel}$ [%]	**EER**$_{known}$ [%]	**EER**$_{novel}$ [%]	**EER**$_{known}$ [%]	**EER**$_{novel}$ [%]
Skin+Latex+EcoFlex	12.4	17.5	16.9	24.6	27.6	34.7	19.0	25.6
Skin+WoodGlue+Latex	13.3	17.8	17.5	24.4	23.0	33.1	17.9	25.1
Skin+Gelatine+Latex	12.7	17.9	17.2	28.8	18.3	29.4	16.1	25.4
Skin+Silgum+Latex	13.2	24.6	14.8	33.2	20.6	31.5	16.2	29.8
Skin+EcoFlex+Silgum	22.5	36.0	23.3	39.4	26.7	43.6	24.2	39.7
Skin+Gelatine+EcoFlex	17.4	27.6	25.0	38.1	22.5	35.8	21.6	33.8
Skin+Silgum+Gelatine	18.1	31.1	23.5	35.5	21.7	27.6	21.1	31.4
Skin+WoodGlue+Silgum	22.7	29.6	23.3	27.6	27.7	42.2	24.6	33.1
Skin+Gelatine+WoodGlue	15.9	22.1	20.2	31.7	18.3	31.1	18.1	28.3
Skin+WoodGlue+EcoFlex	18.8	24.5	21.3	31.3	29.1	30.5	23.1	28.8
Average EER ± STDERROR:	16.7 ± 1.2	**24.9** ± 2.0	20.3 ± 1.1	**31.7** ± 1.7	23.6 ± 1.3	**34.0** ± 1.7	20.2 ± 1.0	**30.1** ± 1.4

DigitalPersona								
	$\mathcal{L}^{BSIF}$		$\mathcal{L}^{LBP}$		$\mathcal{L}^{LPQ}$		**Average**	
Training materials	**EER**$_{known}$ [%]	**EER**$_{novel}$ [%]	**EER**$_{known}$ [%]	**EER**$_{novel}$ [%]	**EER**$_{known}$ [%]	**EER**$_{novel}$ [%]	**EER**$_{known}$ [%]	**EER**$_{novel}$ [%]
Skin+Latex+Playdoh	18.9	16.6	19.9	38.2	3.7	69.4	14.2	41.4
Skin+WoodGlue+Latex	9.4	49.5	19.2	38.5	9.0	50.3	12.5	46.1
Skin+Gelatine+Latex	12.8	19.4	30.4	41.6	7.8	56.3	17.0	39.1
Skin+Silicone+Latex	18.5	22.3	31.4	36.7	12.5	32.1	20.8	30.4
Skin+Playdoh+Silicone	15.3	23.6	18.6	31.9	7.4	32.7	13.8	29.4
Skin+Gelatine+Playdoh	12.1	35.7	18.5	33.1	3.1	66.9	11.2	45.2
Skin+Silicone+Gelatine	13.4	29.4	25.3	40.8	5.7	50.3	14.8	40.2
Skin+WoodGlue+Silicone	12.5	22.9	20.2	32.1	0.6	67.6	11.1	40.9
Skin+Gelatine+WoodGlue	14.1	22.8	22.9	38.2	7.4	45.4	14.8	35.5
Skin+Playdoh+WoodGlue	15.2	17.5	22.1	37.9	9.5	21.6	15.6	25.7
Average EER ± STDERROR:	14.2 ± 0.9	**26.0** ± 3.2	22.9 ± 1.5	**36.9** ± 1.1	6.7 ± 1.1	**49.3** ± 5.2	14.6 ± 0.9	**37.4** ± 2.2

Sagem								
	$\mathcal{L}^{BSIF}$		$\mathcal{L}^{LBP}$		$\mathcal{L}^{LPQ}$		**Average**	
Training materials	**EER**$_{known}$ [%]	**EER**$_{novel}$ [%]	**EER**$_{known}$ [%]	**EER**$_{novel}$ [%]	**EER**$_{known}$ [%]	**EER**$_{novel}$ [%]	**EER**$_{known}$ [%]	**EER**$_{novel}$ [%]
Skin+Latex+Playdoh	11.7	23.4	15.3	23.5	8.7	25.5	11.9	24.1
Skin+WoodGlue+Latex	6.6	50.8	7.1	35.7	3.9	45.4	5.9	44.0
Skin+Gelatine+Latex	12.2	24.2	15.6	16.6	19.6	22.1	15.8	21.0
Skin+Silicone+Latex	11.5	28.6	12.5	36.0	13.5	35.7	12.5	33.4
Skin+Playdoh+Silicone	10.5	29.3	11.3	29.6	10.4	51.6	10.7	36.8
Skin+Gelatine+Playdoh	12.1	39.6	13.9	34.2	12.4	51.3	12.8	41.7
Skin+Silicone+Gelatine	9.3	22.8	10.6	19.9	13.0	51.5	11.0	31.4
Skin+WoodGlue+Silicone	7.0	28.0	10.8	39.5	8.2	34.3	8.7	33.9
Skin+Gelatine+WoodGlue	10.9	25.7	10.8	17.0	17.4	25.9	13.0	22.9
Skin+Playdoh+WoodGlue	8.9	22.6	10.1	24.8	4.5	25.9	7.8	24.4
Average EER ± STDERROR:	10.1 ± 0.7	**29.5** ± 2.9	11.8 ± 0.8	**27.7** ± 2.7	11.2 ± 1.6	**36.9** ± 3.8	11.1 ± 0.9	**31.4** ± 2.6

TABLE III: Average EER, on the test set T_1, of the novel material detector ($\mathcal{M}$) corresponding to various texture descriptors. This is the average over ten different combinations of two materials each that were used for training the novel material detector. The combination of LBP and BGP features resulted in the lowest error rate (which is still very high). In spite of this high error rate, re-training the spoof detector based on the output of the novel material detector improves performance.

Texture descriptors used	**EER**$_{\mathcal{M}}$ $\pm$ **STDERROR** [%]			
	Biometrika	Italdata	Digital Persona	Sagem
Grey Level Co-occurence Matrix (GLCM) [16]	44.6 $\pm$ 1.7	52.3 $\pm$ 2.3	43.7 $\pm$ 2.6	43.6 $\pm$ 3.4
Binary Statistical Image Features (BSIF) [11]	33.2 $\pm$ 1.2	36.9 $\pm$ 1.3	34.2$\pm$ 2.1	38.5$\pm$ 2.7
Local Phase Quantization (LPQ) [13]	34.3 $\pm$ 1.3	36.7 $\pm$ 1.4	44.9 $\pm$ 5.3	40.3 $\pm$ 3.4
Binary Gabor Patterns (BGP) [50]	30.3 $\pm$ 1.0	36.8 $\pm$ 1.4	34.2 $\pm$ 2.3	40.6 $\pm$ 2.2
Local Binary Patterns (LBP) [32]	32.5$\pm$ 2.0	37.3$\pm$1.4	36.6$\pm$ 2.1	31.8$\pm$ 1.7
Local Binary Patterns (LBP) + Binary Gabor Patterns (BGP)	28.5 $\pm$ 1.2	34.1 $\pm$ 1.4	31.1 $\pm$ 2.3	32.5 $\pm$ 2.2

gave the best performance. Thus this configuration was used in subsequent experiments.

Experiment #3: Performance of the spoof detector re-trained on samples flagged by the novel-material detector: The goal of this experiment is to evaluate the performance of the spoof detector when retrained based on the output of the novel-material detector. Tables IV, V, VI and VII show the EER of the retrained spoof detector for the LBP ($\mathcal{L}^{LBP'}$), LPQ ($\mathcal{L}^{LPQ'}$) and BSIF features ($\mathcal{L}^{BSIF'}$), for the Biometrika, Italdata, DigitalPersona and Sagem sensors, respectively. These are the error rates when the spoof detector is retrained using samples that are identified as new spoofs in T_1 and the retrained spoof detector is evaluated on T_2, and vice-versa. In order to detect spoofs made from a novel material, the threshold of the novel material detector was set to the EER point. The novel material detector based on the fused combination LBP+BGP ($\mathcal{M}^{LBP+BGP}$) was used (this combination resulted in the lowest error rate in Experiment #2). Comparison has been made with the performance of the spoof detector ($\mathcal{L}^{LBP}$, $\mathcal{L}^{LPQ}$ and $\mathcal{L}^{BSIF}$) that is not retrained.

The average error *reduction* was 25.3%, 35.2% and 21.0% for the LBP-, LPQ- and BSIF-based spoof detectors, respectively, averaged over T_1 and T_2 and all four sensors, *i.e.*, Biometrika, Italdata, DigitalPersona and Sagem. For instance, EER of the LBP-based spoof detector reduced from 14.0% to 7.7% when retrained using new spoofs from T_2 and evaluated on T_1 for the Biometrika sensor. To highlight this effect, Fig. 6 shows full DET curves reflecting all performance points for the Skin+Silgum+Latex combination from that experiment. Additional curves can be found in the supplemental material. Further, the average error reduction for the LBP-, LPQ- and BSIF-based spoof detectors was 31.4%, 14.8%, 35.5% and 27.1% for the Biometrika, Italdata, DigitalPersona and Sagem sensors, respectively. Furthermore, in comparison to the results reported in [36], average reduction in the error rate of the retrained LPQ and BSIF-based spoof detectors increased by 2.6% and 14.8%, respectively, for the Biometrika sensor. Equivalent performance is reported for the LBP-based spoof detector.

TABLE IV: EER of the **Biometrika**-based spoof detectors retrained using images detected as new spoofs ($\mathcal{L}^{LBP'}, \mathcal{L}^{LPQ'}, \mathcal{L}^{BSIF'}$) in T_1 and evaluated on T_2. Cross-validation is performed by interchanging the role of T_1 and T_2. Comparative assessment has been made with the spoof detector that is not automatically retrained ($\mathcal{L}^{LBP}, \mathcal{L}^{LPQ}, \mathcal{L}^{BSIF}$).

Training materials	Tested on T_2		Tested on T_1	
	$\mathcal{L}^{LBP}$ (not adapted) [%]	$\mathcal{L}^{LBP'}$ (adapted using T_1) [%]	$\mathcal{L}^{LBP}$ (not adapted) [%]	$\mathcal{L}^{LBP'}$ (adapted using T_2) [%]
Skin+Latex+EcoFlex	14.6	13.4	7.0	5.0
Skin+WoodGlue+Latex	12.8	9.6	9.8	6.0
Skin+Gelatine+Latex	13.8	13.4	10.2	7.8
Skin+Silgum+Latex	18.2	14.0	14.2	9.0
Skin+EcoFlex+Silgum	29.6	18.0	21.0	9.0
Skin+Gelatine+EcoFlex	15.2	14.2	10.4	7.2
Skin+Silgum+Gelatine	22.2	15.8	18.2	10.0
Skin+WoodGlue+Silgum	30.4	14.4	27.2	9.2
Skin+Gelatine+WoodGlue	12.2	10.8	10.0	8.2
Skin+WoodGlue+EcoFlex	19.8	12.8	12.2	6.0
Average EER $\pm$ STDERROR :	18.9 $\pm$ 2.1	**13.6** $\pm$ 0.7	14.0 $\pm$ 2.0	**7.7** $\pm$ 0.5

Training materials	Tested on T_2		Tested on T_1	
	$\mathcal{L}^{LPQ}$ (not adapted) [%]	$\mathcal{L}^{LPQ'}$ (adapted using T_1) [%]	$\mathcal{L}^{LPQ}$ (not adapted) [%]	$\mathcal{L}^{LPQ'}$ (adapted using T_2) [%]
Skin+Latex+EcoFlex	19.8	13.0	9.4	6.8
Skin+WoodGlue+Latex	21.6	14.8	9.8	7.6
Skin+Gelatine+Latex	18.6	15.6	10.0	8.8
Skin+Silgum+Latex	18.2	15.8	10.8	9.4
Skin+EcoFlex+Silgum	22.6	16.2	16.2	10.2
Skin+Gelatine+EcoFlex	22.4	16.8	14.2	10.0
Skin+Silgum+Gelatine	20.4	15.6	13.8	10.8
Skin+WoodGlue+Silgum	19.2	15.6	13.6	9.2
Skin+Gelatine+WoodGlue	18.6	12.8	14.0	12.2
Skin+WoodGlue+EcoFlex	22.0	15.4	13.0	9.8
Average EER $\pm$ STDERROR :	20.3 $\pm$ 0.5	**15.2** $\pm$ 0.4	12.5 $\pm$ 0.7	**9.5** $\pm$ 0.5

Training materials	Tested on T_2		Tested on T_1	
	$\mathcal{L}^{BSIF}$ (not adapted) [%]	$\mathcal{L}^{BSIF'}$ (adapted using T_1) [%]	$\mathcal{L}^{BSIF}$ (not adapted) [%]	$\mathcal{L}^{BSIF'}$ (adapted using T_2) [%]
Skin+Latex+EcoFlex	15.8	11.6	7.8	5.4
Skin+WoodGlue+Latex	17.4	12.4	11.2	6.2
Skin+Gelatine+Latex	18.4	14.6	9.4	6.2
Skin+Silgum+Latex	21.4	17.8	11.8	6.8
Skin+EcoFlex+Silgum	28.2	18.6	16.2	9.2
Skin+Gelatine+EcoFlex	24.0	16.0	16.0	10.0
Skin+Silgum+Gelatine	25.2	19.0	15.6	9.8
Skin+WoodGlue+Silgum	25.0	19.4	16.0	8.6
Skin+Gelatine+WoodGlue	19.4	14.0	13.8	9.2
Skin+WoodGlue+EcoFlex	19.8	13.4	13.0	8.2
Average EER $\pm$ STDERROR :	21.5 $\pm$ 1.3	**15.7** $\pm$ 0.9	13.1 $\pm$ 0.9	**8.0** $\pm$ 0.5

Experiment #4: Performance of the spoof detector re-trained using ground-truth (Oracle test):

The goal of this experiment is to evaluate the performance of the spoof detector when retrained using ground-truth. To

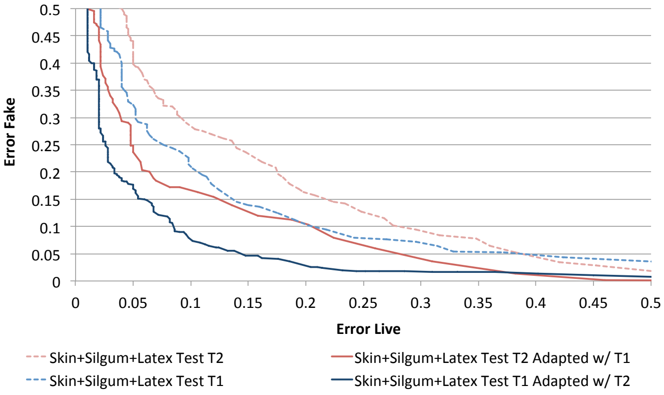

Fig. 6: Full DET curves for the Skin+Silgum+Latex combination for the Biometrika sensor and LBP features from Table IV. In both test cases, the curves shift to the left after adaptation, indicating an improvement in spoof detection performance. Curves for other combinations of materials can be found in the supplemental material.

TABLE V: EER of the **Italdata**-based spoof detectors retrained using images detected as new spoofs ($\mathcal{L}^{LBP'}, \mathcal{L}^{LPQ'}, \mathcal{L}^{BSIF'}$) in T_1 and evaluated on T_2. Cross-validation is performed by interchanging the role of T_1 and T_2. Comparative assessment has been made with the spoof detector that is not automatically retrained ($\mathcal{L}^{LBP}, \mathcal{L}^{LPQ}, \mathcal{L}^{BSIF}$).

Training materials	Tested on T_2		Tested on T_1	
	$\mathcal{L}^{LBP}$ (not adapted) [%]	$\mathcal{L}^{LBP'}$ (adapted using T_1) [%]	$\mathcal{L}^{LBP}$ (not adapted) [%]	$\mathcal{L}^{LBP'}$ (adapted using T_2) [%]
Skin+Latex+EcoFlex	30.3	26.5	26.9	20.4
Skin+WoodGlue+Latex	21.2	25.5	19.7	21.2
Skin+Gelatine+Latex	27.8	24.2	25.3	18.0
Skin+Silgum+Latex	28.8	26.8	26.5	22.0
Skin+EcoFlex+Silgum	31.9	26.8	30.0	23.1
Skin+Gelatine+EcoFlex	37.3	30.0	35.1	25.2
Skin+Silgum+Gelatine	32.2	27.0	29.4	20.1
Skin+WoodGlue+Silgum	36.0	32.6	33.4	30.9
Skin+Gelatine+WoodGlue	27.1	25.8	25.8	20.4
Skin+WoodGlue+EcoFlex	29.0	24.2	31.0	20.0
Average EER ± STDERROR :	30.2 ± 1.5	**26.9** ± 0.8	28.3 ± 1.4	**22.1** ± 1.2

Training materials	Tested on T_2		Tested on T_1	
	$\mathcal{L}^{LPQ}$ (not adapted) [%]	$\mathcal{L}^{LPQ'}$ (adapted using T_1) [%]	$\mathcal{L}^{LPQ}$ (not adapted) [%]	$\mathcal{L}^{LPQ'}$ (adapted using T_2) [%]
Skin+Latex+EcoFlex	21.5	17.0	17.5	17.2
Skin+WoodGlue+Latex	20.6	14.1	20.4	14.9
Skin+Gelatine+Latex	18.6	16.3	14.9	15.1
Skin+Silgum+Latex	24.0	16.1	20.8	15.0
Skin+EcoFlex+Silgum	30.3	25.7	25.0	18.8
Skin+Gelatine+EcoFlex	26.9	21.9	20.3	19.3
Skin+Silgum+Gelatine	22.3	17.0	16.8	17.1
Skin+WoodGlue+Silgum	23.2	18.0	21.7	14.4
Skin+Gelatine+WoodGlue	23.9	16.6	21.1	17.1
Skin+WoodGlue+EcoFlex	24.4	20.2	22.0	16.8
Average EER ± STDERROR :	23.6 ± 1.0	**18.3** ± 1.1	20.1 ± 0.9	**16.6** ± 0.5

Training materials	Tested on T_2		Tested on T_1	
	$\mathcal{L}^{BSIF}$ (not adapted) [%]	$\mathcal{L}^{BSIF'}$ (adapted using T_1) [%]	$\mathcal{L}^{BSIF}$ (not adapted) [%]	$\mathcal{L}^{BSIF'}$ (adapted using T_2) [%]
Skin+Latex+EcoFlex	33.7	25.8	26.3	21.9
Skin+WoodGlue+Latex	23.4	25.6	23.3	29.5
Skin+Gelatine+Latex	26.9	25.1	22.2	22.2
Skin+Silgum+Latex	27.8	27.7	27.7	26.5
Skin+EcoFlex+Silgum	37.1	31.4	32.9	27.6
Skin+Gelatine+EcoFlex	39.7	31.2	33.7	27.3
Skin+Silgum+Gelatine	34.6	29.7	28.9	25.8
Skin+WoodGlue+Silgum	35.0	32.0	34.7	35.3
Skin+Gelatine+WoodGlue	28.1	27.1	23.9	26.2
Skin+WoodGlue+EcoFlex	29.0	25.4	27.6	23.6
Average EER ± STDERROR :	31.5 ± 1.6	**28.1** ± 0.9	28.1 ± 1.4	**26.6** ± 1.2

TABLE VI: EER of the **DigitalPersona**-based spoof detectors retrained using images detected as new spoofs ($\mathcal{L}^{LBP'}, \mathcal{L}^{LPQ'}, \mathcal{L}^{BSIF'}$) in T_1 and evaluated on T_2. Cross-validation is performed by interchanging the role of T_1 and T_2. Comparative assessment has been made with the spoof detector that is not automatically retrained ($\mathcal{L}^{LBP}, \mathcal{L}^{LPQ}, \mathcal{L}^{BSIF}$).

Training materials	Tested on T_2		Tested on T_1	
	$\mathcal{L}^{LBP}$ (not adapted) [%]	$\mathcal{L}^{LBP'}$ (adapted using T_1) [%]	$\mathcal{L}^{LBP}$ (not adapted) [%]	$\mathcal{L}^{LBP'}$ (adapted using T_2) [%]
Skin+Latex+Playdoh	38.9	30.6	32.4	23.4
Skin+WoodGlue+Latex	37.7	28.1	39.1	27.7
Skin+Gelatine+Latex	34.2	21.6	42.9	28.9
Skin+Silicone+Latex	44.0	24.6	47.5	24.8
Skin+Playdoh+Silicone	39.6	29.4	36.5	20.6
Skin+Gelatine+Playdoh	27.4	23.6	30.9	21.9
Skin+Silicone+Gelatine	38.0	30.9	48.7	39.5
Skin+WoodGlue+Silicone	38.3	23.6	40.4	21.3
Skin+Gelatine+WoodGlue	28.8	20.6	36.8	24.6
Skin+Playdoh+WoodGlue	36.9	32.9	30.9	23.0
Average EER ± STDERROR :	36.4 ± 1.6	**26.6** ± 1.4	38.6 ± 2.0	**25.6** ± 1.8

Training materials	Tested on T_2		Tested on T_1	
	$\mathcal{L}^{LPQ}$ (not adapted) [%]	$\mathcal{L}^{LPQ'}$ (adapted using T_1) [%]	$\mathcal{L}^{LPQ}$ (not adapted) [%]	$\mathcal{L}^{LPQ'}$ (adapted using T_2) [%]
Skin+Latex+Playdoh	44.1	17.4	43.8	10.1
Skin+WoodGlue+Latex	42.9	19.0	37.1	11.9
Skin+Gelatine+Latex	33.2	17.1	32.4	10.0
Skin+Silicone+Latex	42.6	18.9	36.0	8.1
Skin+Playdoh+Silicone	27.1	18.9	26.3	8.8
Skin+Gelatine+Playdoh	50.1	23.8	54.0	15.4
Skin+Silicone+Gelatine	38.5	24.1	44.5	14.1
Skin+WoodGlue+Silicone	47.8	20.3	42.4	9.0
Skin+Gelatine+WoodGlue	38.8	20.5	37.8	11.2
Skin+Playdoh+WoodGlue	32.5	16.4	25.4	9.5
Average EER ± STDERROR :	39.8 ± 2.3	**19.6** ± 0.8	38.0 ± 2.8	**10.8** ± 0.8

Training materials	Tested on T_2		Tested on T_1	
	$\mathcal{L}^{BSIF}$ (not adapted) [%]	$\mathcal{L}^{BSIF'}$ (adapted using T_1) [%]	$\mathcal{L}^{BSIF}$ (not adapted) [%]	$\mathcal{L}^{BSIF'}$ (adapted using T_2) [%]
Skin+Latex+Playdoh	26.4	22.4	22.7	18.0
Skin+WoodGlue+Latex	36.9	32.0	32.6	26.6
Skin+Gelatine+Latex	21.9	20.4	21.5	18.1
Skin+Silicone+Latex	29.5	24.3	23.5	18.2
Skin+Playdoh+Silicone	24.6	21.3	21.0	16.5
Skin+Gelatine+Playdoh	29.3	29.4	30.0	21.8
Skin+Silicone+Gelatine	26.1	23.1	27.5	19.8
Skin+WoodGlue+Silicone	27.7	21.2	21.8	18.0
Skin+Gelatine+WoodGlue	21.8	21.8	21.6	19.7
Skin+Playdoh+WoodGlue	23.0	22.4	19.3	18.7
Average EER ± STDERROR :	26.7 ± 1.4	**23.8** ± 1.2	24.2 ± 1.4	**19.5** ± 0.9

this aim, the LBP-, LPQ- and BSIF- based spoof detectors are retrained using all the samples in test sets together with the class label.

Table VIII show the EER of the oracle test for the LBP ($\mathcal{L}^{LBP'}$), LPQ ($\mathcal{L}^{LPQ'}$) and BSIF features ($\mathcal{L}^{BSIF'}$), for Biometrika, Italdata, DigitalPersona and Sagem sensors, respectively. These are the error rates when the spoof detector is retrained using samples that are identified as new spoofs in T_1 and the retrained spoof detector is evaluated on T_2, and vice-versa. Comparison has been made with the performance of the spoof detector automatically retrained using the output of the novel-material detector (see Experiment #3).

The average error *reduction* was 33.5%, 40.6% and 27.4% for the LBP-, LPQ- and BSIF-based spoof detectors retrained using ground-truth, respectively, averaged over T_1 and T_2 and all four sensors, *i.e.*, Biometrika, Italdata, DigitalPersona and Sagem. The difference in the average error reduction in comparison to Experiment #3 was 8.2%, 5.4% and 6.5%.

Further, the average error reduction for the LBP-, LPQ- and BSIF-based spoof detectors was 34.1%, 21.2%, 43.8% and 36.5% for the Biometrika, Italdata, DigitalPersona and

TABLE VII: EER of the **Sagem**-based spoof detectors retrained using images detected as new spoofs ($\mathcal{L}^{LBP'}, \mathcal{L}^{LPQ'}, \mathcal{L}^{BSIF'}$) in T_1 and evaluated on T_2. Cross-validation is performed by interchanging the role of T_1 and T_2. Comparative assessment has been made with the spoof detector that is not automatically retrained ($\mathcal{L}^{LBP}, \mathcal{L}^{LPQ}, \mathcal{L}^{BSIF}$).

Training materials	Tested on T_2		Tested on T_1	
	$\mathcal{L}^{LBP}$	$\mathcal{L}^{LBP'}$	$\mathcal{L}^{LBP}$	$\mathcal{L}^{LBP'}$
	(not adapted) [%]	(adapted using T_1) [%]	(not adapted) [%]	(adapted using T_2) [%]
Skin+Latex+Playdoh	21.8	17.9	18.0	17.9
Skin+WoodGlue+Latex	30.6	19.4	23.1	21.0
Skin+Gelatine+Latex	14.8	13.0	15.8	16.5
Skin+Silicone+Latex	26.7	19.9	22.7	21.0
Skin+Playdoh+Silicone	28.6	21.9	21.1	16.3
Skin+Gelatine+Playdoh	21.9	18.6	30.2	20.2
Skin+Silicone+Gelatine	14.5	12.4	19.3	16.6
Skin+WoodGlue+Silicone	31.4	16.5	25.4	19.5
Skin+Gelatine+WoodGlue	13.9	12.7	15.6	15.3
Skin+Playdoh+WoodGlue	21.9	19.3	18.5	16.8
Average EER ± STDERROR :	22.6 ± 2.1	**17.2** ± 1.1	21.0 ± 1.4	**18.1** ± 0.7

Training materials	Tested on T_2		Tested on T_1	
	$\mathcal{L}^{LPQ}$	$\mathcal{L}^{LPQ'}$	$\mathcal{L}^{LPQ}$	$\mathcal{L}^{LPQ'}$
	(not adapted) [%]	(adapted using T_1) [%]	(not adapted) [%]	(adapted using T_2) [%]
Skin+Latex+Playdoh	17.9	15.9	20.2	16.5
Skin+WoodGlue+Latex	39.2	19.8	30.6	17.9
Skin+Gelatine+Latex	19.2	12.0	16.6	15.2
Skin+Silicone+Latex	28.0	19.6	21.8	18.3
Skin+Playdoh+Silicone	34.3	20.4	35.7	16.7
Skin+Gelatine+Playdoh	32.3	16.8	36.1	20.3
Skin+Silicone+Gelatine	34.0	15.0	31.4	17.3
Skin+WoodGlue+Silicone	28.9	19.5	22.5	16.6
Skin+Gelatine+WoodGlue	20.9	13.5	19.6	16.1
Skin+Playdoh+WoodGlue	17.8	16.9	19.3	16.4
Average EER ± STDERROR :	27.3 ± 2.5	**16.9** ± 0.9	25.4 ± 2.3	**17.1** ± 0.4

Training materials	Tested on T_2		Tested on T_1	
	$\mathcal{L}^{BSIF}$	$\mathcal{L}^{BSIF'}$	$\mathcal{L}^{BSIF}$	$\mathcal{L}^{BSIF'}$
	(not adapted) [%]	(adapted using T_1) [%]	(not adapted) [%]	(adapted using T_2) [%]
Skin+Latex+Playdoh	20.7	14.3	21.4	20.0
Skin+WoodGlue+Latex	46.6	18.1	39.4	19.1
Skin+Gelatine+Latex	18.5	12.7	22.3	19.6
Skin+Silicone+Latex	24.7	18.2	22.7	20.0
Skin+Playdoh+Silicone	25.0	21.1	29.6	21.9
Skin+Gelatine+Playdoh	23.6	20.6	35.0	27.2
Skin+Silicone+Gelatine	21.8	14.2	26.3	18.5
Skin+WoodGlue+Silicone	23.2	16.9	20.4	18.2
Skin+Gelatine+WoodGlue	18.2	12.0	25.8	21.2
Skin+Playdoh+WoodGlue	19.5	14.3	23.9	21.1
Average EER ± STDERROR :	24.2 ± 2.6	**16.2** ± 1.0	26.7 ± 2.0	**20.7** ± 0.8

TABLE VIII: Average EER of the **Biometrika**, **Italdata**, **DigitalPersona**, **Sagem**-based spoof detector retrained using the ground-truth ($\mathcal{L}^{LBP'}, \mathcal{L}^{LPQ'}, \mathcal{L}^{BSIF'}$), *i.e.*, retrained using all of the samples in T_1 and evaluated on T_2. Cross-validation is performed by interchanging the role of T_1 and T_2. Comparative assessment has been made with the spoof detector that is not automatically retrained ($\mathcal{L}^{LBP}, \mathcal{L}^{LPQ}, \mathcal{L}^{BSIF}$). Tables with the complete data used to build this summary can be found in the supplemental material, along with full DET curves for the Biometrika + LBP combination.

Sensors	Tested on T_2		Tested on T_1	
	(not adapted) [%]	(adapted using T_1) [%]	(not adapted) [%]	(adapted using T_2) [%]
Biometrika				
	$\mathcal{L}^{LBP}$	$\mathcal{L}^{LBP'}$	$\mathcal{L}^{LBP}$	$\mathcal{L}^{LBP'}$
Average EER STDERROR :	18.9 ± 2.1	**13.5** ± 0.6	14.0 ± 2.0	**7.7** ± 0.4
	$\mathcal{L}^{LPQ}$	$\mathcal{L}^{LPQ'}$	$\mathcal{L}^{LPQ}$	$\mathcal{L}^{LPQ'}$
Average EER ± STDERROR:	20.3 ± 0.5	**14.6** ± 0.5	12.5 ± 0.7	**9.0** ± 0.5
	$\mathcal{L}^{BSIF}$	$\mathcal{L}^{BSIF'}$	$\mathcal{L}^{BSIF}$	$\mathcal{L}^{BSIF'}$
Average EER ± STDERROR:	21.5 ± 1.3	**15.4** ± 0.6	13.1 ± 0.9	**7.0** ± 0.4
Italdata				
	$\mathcal{L}^{LBP}$	$\mathcal{L}^{LBP'}$	$\mathcal{L}^{LBP}$	$\mathcal{L}^{LBP'}$
Average EER ± STDERROR:	30.2 ± 1.5	**24.6** ± 0.4	28.3 ± 1.4	**18.9** ± 0.7
	$\mathcal{L}^{LPQ}$	$\mathcal{L}^{LPQ'}$	$\mathcal{L}^{LPQ}$	$\mathcal{L}^{LPQ'}$
Average EER ± STDERROR :	23.6 ± 1.0	**17.6** ± 0.9	20.0 ± 0.9	**15.8** ± 0.6
	$\mathcal{L}^{BSIF}$	$\mathcal{L}^{BSIF'}$	$\mathcal{L}^{BSIF}$	$\mathcal{L}^{BSIF'}$
Average EER ± STDERROR :	31.5 ± 1.6	**26.7** ± 0.6	28.1 ± 1.4	**24.3** ± 1.0
DigitalPersona				
	$\mathcal{L}^{LBP}$	$\mathcal{L}^{LBP'}$	$\mathcal{L}^{LBP}$	$\mathcal{L}^{LBP'}$
Average EER ± STDERROR :	36.4 ± 1.7	**22.1** ± 0.9	38.6 ± 2.1	**23.6** ± 1.5
	$\mathcal{L}^{LPQ}$	$\mathcal{L}^{LPQ'}$	$\mathcal{L}^{LPQ}$	$\mathcal{L}^{LPQ'}$
Average EER ± STDERROR :	39.8 ± 2.3	**14.8** ± 0.3	38.0 ± 2.8	**9.2** ± 0.3
	$\mathcal{L}^{BSIF}$	$\mathcal{L}^{BSIF'}$	$\mathcal{L}^{BSIF}$	$\mathcal{L}^{BSIF'}$
Average EER ± STDERROR :	26.7 ± 1.4	**21.2** ± 0.2	24.2 ± 1.4	**18.1** ± 0.3
Sagem				
	$\mathcal{L}^{LBP}$	$\mathcal{L}^{LBP'}$	$\mathcal{L}^{LBP}$	$\mathcal{L}^{LBP'}$
Average EER ± STDERROR :	22.6 ± 2.2	**14.6** ± 0.5	21.0 ± 1.5	**14.8** ± 0.3
	$\mathcal{L}^{LPQ}$	$\mathcal{L}^{LPQ'}$	$\mathcal{L}^{LPQ}$	$\mathcal{L}^{LPQ'}$
Average EER ± STDERROR :	27.3 ± 2.5	**14.6** ± 0.7	25.4 ± 2.3	**15.9** ± 0.6
	$\mathcal{L}^{BSIF}$	$\mathcal{L}^{BSIF'}$	$\mathcal{L}^{BSIF}$	$\mathcal{L}^{BSIF'}$
Average EER ± STDERROR :	24.2 ± 2.6	**14.7** ± 0.4	26.7 ± 2.0	**18.5** ± 0.6

Sagem sensors, respectively. The difference in the average error reduction in comparison to Experiment #3 was 2.7%, 6.4%, 8.3% and 9.4%. The marginal difference in the error reduction between the spoof detector retrained using ground-truth and output of the novel material detector shows the efficacy of the proposed automatic adaptation scheme based on the W-SVM while indicating that more research is needed on this topic.

VII. Conclusion

Recent studies suggest a threefold increase in the error rate of a fingerprint spoof detector when encountering spoofs generated using materials that were not observed during the training stage. In this article, we proposed a scheme for the automatic detection and adaptation of a spoof detector to spoofs fabricated using novel materials that are encountered during the operational phase, thus addressing the underlying open set recognition problem. To this end, a W-SVM-based [39] novel-material detector was developed to detect spoofs made of new materials. Detected samples are used to *automatically* retrain and update a W-SVM-based spoof detector. The performance of the fingerprint spoof detector when retrained based on the output of the novel material detector improves by up to 44%. As the proposed adaptation scheme is offline, any retraining overhead does not affect the throughput of the spoof detector.

One weakness of this approach is the potential to misclassify lower quality live prints as novel spoofs. The cumulative effect of retraining the spoof detector using live samples, which are incorrectly detected as new spoofs by the novel material detector, will be counter-productive over a period of time. A future direction that may prove to be fruitful is feature learning [3], whereby strongly invariant representations can be learned for each known class, thus improving the detection rate of low-quality live samples. Coupled with approaches that are tailored to open set recognition like the W-SVM, this could be a powerful solution to the overall problem.

References

[1] A. Abhyankar and S. Schuckers. Integrating a wavelet based perspiration liveness check with fingerprint recognition. *Pattern Recognition*, 42:452 – 464, 2009.

[2] S. S. Arora, K. Cao, A. K. Jain, and N.G. Paulter Jr. 3d fingerprint phantoms. Technical report, Michigan State University, East Lansing, Department of Computer Science and Engineering, 2013.

[3] Y. Bengio. *Learning Deep Architectures for AI*. Now Publishers, 2009.

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

[5] C.-C. Chang and C.-J. Lin. LIBSVM: A library for support vector machines. *ACM Transactions on Intelligent Systems and Technology*, 2:27:1–27:27, 2011.

[6] S. Coles. *An Introduction to Statistical Modeling of Extreme Values*. Springer, 2001.

[7] P. Coli, G.L. Marcialis, and F. Roli. Power spectrum-based fingerprint vitality detection. In *Proc. of IEEE Intl. Work. on Automatic Identification Advanced Technologies AutoID*, pages 169–173, Alghero, Italy, 2007.

[8] M. V. de Water. Can fingerprints be forged. *The Science News-Letter*, 29:90–92, 1936.

[9] M. Espinoza and C. Champod. Using the number of pores on fingerprint images to detect spoofing attacks. In *Proc. of Intl. Conf. on Hand-based Biometrics*, page 15, Hong Kong, China, 201.

[10] L. Ghiani, P. Denti, and G. L. Marcialis. Experimental results on fingerprint liveness detection. In *Proc. of IEEE Intl. Conf. on Articulated Motion and Deformable Objects*, pages 210–218, Spain, 2012.

[11] L. Ghiani, A. Hadid, G.L. Marcialis, and F. Roli. Fingerprint liveness detection using binarized statistical image features. In *Proc. of IEEE Intl Conf. on Biometrics: Theory, Applications and Systems*, pages 1–6, Sept 2013.

[12] L. Ghiani, G. L. Marcialis, and F. Roli. Experimental results on the feature-level fusion of multiple fingerprint liveness detection algorithms. In *Proc. of ACM Workshop on Multimedia and Security*, pages 157–164, Coventry, UK, 2012.

[13] L. Ghiani, G.L. Marcialis, and F. Roli. Fingerprint liveness detection by local phase quantization. In *Proc. of Intl. Conf. On Pattern Recognition*, pages 537–540, 2012.

[14] L. Ghiani, D. Yambay, V. Mura, S. Tocco, G.L. Marcialis, F. Roli, and S. Schuckers. LivDet 2013 Fingerprint Liveness Detection Competition 2013. http://prag.diee.unica.it/fldc-tset/LivDet_2013.pdf, 2013.

[15] D. Gragnaniello, G. Poggi, C. Sansone, and L. Verdoliva. Fingerprint liveness detection based on weber local image descriptor. In *Proc. of IEEE Workshop on Biometric Measurements and Systems for Security and Medical Applications*, pages 46–50, Sept 2013.

[16] R.M. Haralick, K. Shanmugan, and I. Dinstein. Textural features for image classification. *IEEE Trans. on Systems, Man, and Cybernetics*, 3:610–621, 1973.

[17] L.P. Jain, W.J. Scheirer, and T.E. Boult. Multi-class open set recognition using probability of inclusion. In *The European Conference on Computer Vision (ECCV)*, September 2014.

[18] X. Jia, X. Yang, Y. Zang, N. Zhang, R. Dai, J. Tian, and J. Zhao. Multi-scale block local ternary patterns for fingerprints vitality detection. In *Proc. of Intl. Conf. on Biometrics*, pages 1–6, June 2013.

[19] H. Kang, B. Lee, H. Kim, D. Shin, and J. Kim. A study on performance evaluation of the liveness detection for various fingerprint sensor modules. In *Lecture Notes in Computer Science*, pages 1245–1253, 2003.

[20] M. Manevitz and M. Yousef. One-class SVMs for document classification. *JMLR*, 2:139–154, March 2002.

[21] E. Marasco and A. Ross. A survey on anti-spoofing schemes for fingerprints. *ACM Computing Surveys*, pages 1–35, 2014.

[22] E. Marasco and C. Sansone. On the robustness of fingerprint liveness detection algorithms against new materials used for spoofing. In *Proc. of Intl. Conf. on Bio-Inspired Systems and Signal Processing*, pages 553–558, Rome, Italy, 2011.

[23] E. Marasco and C. Sansone. Combining perspiration- and morphology-based static features for fingerprint liveness detection. *Pattern Recognition Letters*, 33:1148–1156, 2012.

[24] S. Marcel, M. Nixon, and S. Z. Li. *Handbook of Biometric Anti-Spoofing*. Springer, 2014.

[25] G.L. Marcialis, F. Roli, and A. Tidu. Analysis of fingerprint pores for vitality detection. In *Proc. of Intl. Conf. on Pattern Recognition (ICPR)*, pages 1289–1292, Aug 2010.

[26] T. Matsumoto, H. Matsumoto, K. Yamada, and S. Hoshino. Impact of artificial "gummy" fingers on fingerprint systems. In *Proc. of SPIE Opt. Sec. Counterfeit Deterrence Tech. IV*, pages 275–289, 2002.

[27] Y.S. Moon, J.S. Chen, K.C. Chan, K. So, and K.S. Woo. Wavelet based fingerprint liveness detection. *Electronic Letters*, 41:1112–1113, 2005.

[28] A. Niculescu-Mizil and R. Caruana. Predicting good probabilities with supervised learning. In *ICML*, pages 625–632, 2005.

[29] S.B. Nikam and S. Aggarwal. Local binary pattern and wavelet-based spoof fingerprint detection. *Intl. Journal of Biometrics*, 1(2):141–159, 2008.

[30] S.B. Nikam and S. Aggarwal. Wavelet energy signature and glcm features-based fingerprint anti-spoofing. In *Proc. of IEEE Int. Conf. On Wavelet Analysis and Pattern Recognition*, pages 717 –723, Hong Kong, China, 2008.

[31] K. Nixon, V. Aimale, and R.K. Rowe. Spoof detection schemes. In A.K. Jain, P. Flynn, and A. Ross, editors, *Handbook of Biometrics*, pages 403–423. Springer US, 2008.

[32] T. Ojala, M. Pietikinen, and T. Menp. Multiresolution gray-scale and rotation invariant texture classification with local binary patterns. *IEEE Trans. on Pattern Analysis and Machine Intelligence*, 24(7):971–987, 2002.

[33] J. Platt. Probabilistic outputs for support vector machines and comparison to regularized likelihood methods. In A. Smola, P. Bartlett, and B. Schölkopf, editors, *Advances in Large Margin Classifiers*. MIT Press, 2000.

[34] A. Rattani and N. Poh. Biometric system design under zero and non-zero effort attacks. In *Proc. of IEEE Intl. Conf. on Biometrics*, Madrid, Spain, 2013.

[35] A. Rattani, N. Poh, and A. Ross. A bayesian approach for modeling sensor influence on quality, liveness and match score values in fingerprint verification. In *Proc. of IEEE Intl. Workshop on Information Forensics and Security*, pages 1–6, Guangzhou, China, 2013.

[36] A. Rattani and A. Ross. Automatic adaptation of fingerprint liveness detector to new spoof materials. In *Proc. of IEEE Intl. Joint Conf. on Biometrics*, Clearwater, Florida, 2014.

[37] A. Rattani and A. Ross. Minimizing the impact of spoof fabrication material on fingerprint liveness detector. In *Proc. of IEEE Intl. Conf. on Image Processing*, Paris, France, 2014.

[38] A. Ross, J. Shah, and A. K. Jain. From template to image: Reconstructing fingerprints from minutiae points. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 29:544–560, 2007.

[39] W.J. Scheirer, L.P. Jain, and T.E. Boult. Probability models for open set recognition. *IEEE Transactions on Pattern Analysis and Machine Intelligence (T-PAMI)*, 36, November 2014.

[40] W.J. Scheirer, N. Kumar, P.N. Belhumeur, and T.E. Boult. Multi-attribute spaces: Calibration for attribute fusion and similarity search. In *IEEE CVPR*, June 2012.

[41] W.J. Scheirer, A. Rocha, R. Michaels, and T.E. Boult. Meta-recognition: The theory and practice of recognition score analysis. *IEEE T-PAMI*, 33(8):1689–1695, August 2011.

[42] W.J. Scheirer, A. Rocha, A. Sapkota, and T.E. Boult. Towards open set recognition. *IEEE Transactions on Pattern Analysis and Machine Intelligence (T-PAMI)*, 36:1757–1772, July 2013.

[43] B. Schölkopf, J.C. Platt, J.C. Shawe-Taylor, A.J. Smola, and R.C. Williamson. Estimating the support of a high-dimensional distribution. *Neural Computation*, 13(7):1443–1471, July 2001.

[44] B. Tan, A. Lewicke, D. Yambay, and S. Schuckers. The effect of environmental conditions and novel spoofing methods on fingerprint anti-spoofing algorithms. In *Proc. of IEEE Intl. Workshop on Information Forensics and Security*, pages 1–6, 2010.

[45] B. Tan and S. Schuckers. Liveness detection for fingerprint scanners based on the statistics of wavelet signal processing. In *Proc. of Workshop on Biometrics in Computer Vision and Pattern Recognition*, pages 26–26, June 2006.

[46] B. Tan and S. Schuckers. Spoofing protection for fingerprint scanner by fusing ridge signal and valley noise. *Pattern Recognition*, 43(8):2845–2857, 2010.

[47] A. Wehde and J. N. Beffel. Fingerprints can be forged. *Tremonia Publish Co.*, 1924.

[48] D. Yambay, L. Ghiani, P. Denti, G. L. Marcialis, F. Roli, and S. Schuckers. LivDet 2011 - fingerprint liveness detection competition. In *Proc. of IEEE Intl. Conf. on Biometrics*, pages 208–215, Delhi, India, 2012.

[49] B. Zadrozny and C. Elkan. Transforming classifier scores into accurate multiclass probability estimates. In *Proceedings of the Eighth ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, pages 694–699, 2002.

[50] L. Zhang, Z. Zhou, and H. Li. Binary gabor pattern: An efficient and robust descriptor for texture classification. In *Proc. of IEEE Intl. Conf. on Image Processing*, FL, USA, 2012.

[51] X. Zhou and T. Huang. Relevance feedback in image retrieval: A comprehensive review. *Multimedia Systems*, 8(6):536–544, 2003.