A Realistic Evaluation of Iris Presentation Attack Detection

Ana F. Sequeira[1], Shejin Thavalengal[2], James Ferryman[1], Peter Corcoran[2], and Jaime S. Cardoso[3]

[1]School of Systems Engineering, University of Reading, UK.
[2]C3 Imaging Centre, National University of Ireland, Galway, Ireland.
[3]FEUP, University of Porto and INESC TEC, Porto, Portugal.
Email: {a.f.p.sequeira, j.m.ferryman}@reading.ac.uk, {s.thavalengal1, peter.corcoran}@nuigalway.ie, jsc@fe.up.pt

***Abstract*—Iris liveness detection methods have been developed to overcome the vulnerability of iris biometric systems to spoofing attacks. In the literature, it is typically assumed that a known attack modality will be perpetrated. Then liveness models are designed using labelled samples from both real/live and fake/spoof distributions, the latter derived from the assumed attack modality. In this work it is argued that a comprehensive modelling of the spoof samples is not possible in a real-world scenario where the attack modality cannot be known with a high degree of certainty. In fact making this assumption will render the liveness detection system more vulnerable to attacks that were not included in the original training. To provide a more realistic evaluation, this work proposes: a) testing the binary models with unknown spoof samples that were not present in the training step; b) the use of a single-class classification designing the classifier by modelling only the distribution of live samples. The results obtained support the assertion that many evaluation methods from the literature are misleading and may lead to optimistic estimates of the robustness of liveness detection in practical use cases.**

***Keywords*—biometrics; iris; presentation attack; one-class classification**

I. Introduction

Biometric recognition is nowadays a mature technology used in many government and civilian applications such as e-passports, ID cards, and border control. While faces and fingerprints are the best known biometrics, the use of iris for authentication has grown as it is more difficult to spoof than faces recognition or fingerprint based systems [1].

In most use cases, the biometric credentials are acquired in a controlled environment and under supervision. Thus there is a low risk of the user providing false credential or attempting to spoof the system. However over the last decade there has been a wide roll-out of unconstrained acquisition systems- the use of iris biometrics in many airports is a well known example [2]. In these cases the system must be more flexible in its tolerances and as these systems are less closely supervised there is greater scope for tampering and spoofing the system.

More recently, iris is being considered for unsupervised applications, in particular the use of iris as a secure authentication mechanism for smartphones is imminent [3], [4].

Biometric recognition systems in general, and iris recognition systems in particular, can be spoofed by presenting fake or altered samples of the biometric trait at the sensor [5]. These spoofing/presentation attacks can include printed images, patterned contact lenses, videos or images displayed in electronic devices (referred as “Electronic Display/Screen Attacks” [6]). Liveness detection techniques and tamper detection methods are considered as *presentation attack detection* (PAD) methods [7] and are intended to detect spoofing attacks.

A significant body of literature on PAD methods is available [6], [8]. It can be argued here that a majority of these techniques are based on evaluation methodologies that are flawed. More specifically, the classification models employed in most PAD methods are designed using datasets from live samples and a specific type of spoof samples. Then evaluation is based on the use of samples provenient of the same populations. This work argues that this approach to performance evaluation of liveness methods is overly optimistic as it does not consider scenarios where a spoof sample is significantly different from the spoof samples used for training. It may well happen that such a sample has a higher probability to circumvent the system than samples drawn from the original training dataset.

To overcome this problem, it is important to test the PAD system with additional spoof samples taken from different sources than the original training dataset. This reduces the dependencies on the original datasets and should give a more honest appraisal of system vulnerability. However it is always tempting to include the “unseen” spoof data into the training process so there are limitations to this approach as well. Thus we propose and implement a liveness detection system based on single-class training. Here the classifier design is based on modelling the distribution of live samples only - spoof data is not used. The system predicts presentation attacks based on samples that are unlikely to fall within the designed single-class. The expected performance will be lower for “known” attacks, but the real purpose of this preliminary work is to show that it can be more robust against “unseen” attacks from outside the original training dataset.

The remaining sections of this paper are organised as follows. In section II methodological limitations of the current research are pointed out. Also, the proposed PAD method is

Ana F. Sequeira is funded by the EU FASTPASS project under Grant agreement 312583. Shejin Thavalengal is funded by the employment based PhD program of the Irish Research Council.

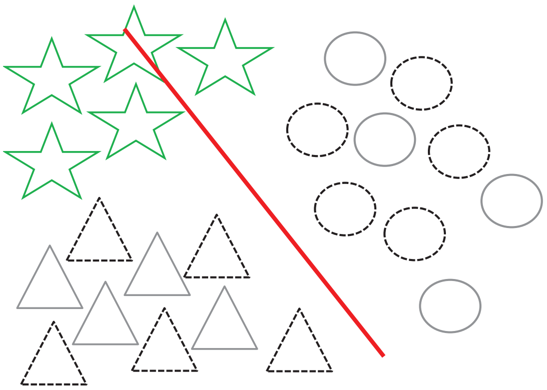

Fig. 1. Illustrative example of the mismatch between training and testing data when a binary classifier is used. Black dotted triangles and circles are spoof and live samples in the training set. Grey circles are live samples in the test set; and grey triangles are spoof samples in the test set similar to fake samples present in the training set; green stars are spoof samples from an unseen attack presented in test step. The red curve represents the binary classification boundary learnt from the training samples.

discussed in this section. In section III the experimental setup is presented. The results and their discussion are presented in section IV and the article is concluded in section V.

II. Iris PAD methods: Limitations and Proposed Approaches

A. Limitations of current approaches

A methodological limitation of the majority of current approaches is that these techniques are designed and evaluated using spoof samples drawn from a limited dataset of spoof samples. Often these samples originate from a single attack modality. The systems are developed and tested under the assumption that the intruder will perpetrate the same spoofing attack. This results in an optimistic estimation of the security level of the system. At the design time, the developer presumes to possess labelled data representative of the live and spoof samples and therefore commonly employs standard binary classification as a detection tool [6], [9], [10].

The binary classifiers used to make the decision between live and spoof samples implicitly assume that the training samples are representative of the complete population. Furthermore it is common practice to draw test data from the same distribution as the training data in order to evaluate the underlying detection system. Although that might be a fair assumption for the live samples, it can be argued that it may be a crude model for spoof samples obtained from different attack modalities. As depicted in Fig. 1, it may happen that there is a mismatch between the distribution of observations in training and testing data. With new applications, particularly unconstrained and unsupervised use cases as exemplified by smartphone biometrics, there is a need to consider a wider range of attack modalities.

The pioneer works that raised the question of evaluating liveness detection methods across different types of spoofing samples appeared in the fingerprint domain. Marasco and Sansone [11] performed an experimental comparison of fingerprint liveness detection approaches adopting materials for training different than those adopted for testing. This experiment allowed to conclude that the performance significantly decreases in such conditions. Rattani and Ross [12], designed an online scheme for automatic adaptation of a liveness detector to novel spoof materials encountered during the operational phase. This can lead to significant improvements in the performance. Although this method well accommodates small differences between samples, it will likely under-perform when the new samples are significantly different. The latter work was expanded and a single classifier (Weibull-calibrated SVM) was proposed to perform both the novel material detection and the spoof detection [13]. Sequeira and Cardoso [14] compared traditional fingerprint liveness detection methodologies with approaches that incorporate the evaluation of binary classification models with unseen materials. The design of models for one-class classification relying only on the information of real samples is also investigated in this work.

The question raised in this work is pertinent regardless of the biometric trait used. For most biometric traits there are several different possibilities for building fake samples. Surprisingly, regarding iris liveness, the works in the literature do not consider scenarios where the testing samples are derived from different types of presentation attacks than those represented in the original training dataset.

The single exception is the work presented by Bowyer and Doyle [15] concerning spoofing attacks using contact lenses. The authors perform a baseline experiment where the train and test datasets each contained iris images with three lens types. Using the same texture features and classifiers, the authors repeated the experiment with a training set with two of the three lens types and the test set with the third lens type. From the baseline experiment to the new one the correct classification error rate lowered from 100% to a worst case of 75%. These results illustrate how experimental results obtained using the same lens types in both the training and testing data can give a very misleading idea of the accuracy that will result when a new lens type is encountered.

Though Bowyer's approach is novel on evaluating unseen spoof samples, it relies on a binary classifier and comprises only one type of attack. In the literature, works can be found combining methodologies for different types of spoofing attacks. For example, features designed for print attack and contact lenses attack [16]. Nevertheless the evaluation considering different types of spoof samples was limited by the fact that existing databases for iris liveness detection contain fake samples produced by one single type of attack. The recent construction of a new database comprising several types of iris spoofing attacks [6] allows new evaluation scenarios.

B. Realistic evaluation of iris PAD methods

As new types of spoofing attacks continuously appear and become more sophisticated, a fair evaluation of methods should not rely only on previously known attacks. The path

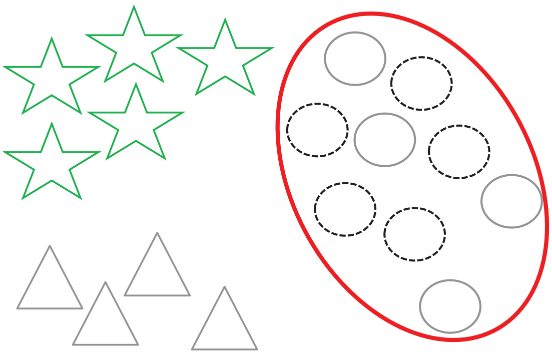

Fig. 2. Illustrative example of how an one-class classifier accomodates the presence of unseen attacks in test step. Black dotted circles are live samples in the training set. Grey circles are live samples in the test set. Grey triangles are spoof samples in the test set; green stars are spoof samples from an unseen attack presented in test step. The red curve represents the one-class model learnt from the training live samples.

to pursue is to use less information about the spoof samples and, in the limit, rely strongly on the information about the real samples. Such an approach may better accommodate new unseen attacks that were not present in the design phase of the model. This is depicted in Fig. 2.

The aim of this work is to evaluate PAD methods realistically by considering classification scenarios where the test data is not exclusively drawn from the same distribution as the training data. Evaluating the PAD models with data for testing coming from the same distribution of the training data, may be a fair assumption for the live samples and extensively used in the literature [6], [8], [9], [10]. But, it may happen to be an inexact model for the spoof samples considering that spoofing techniques are evolving [1]. Therefore, the study presented compares the traditional approach with two different scenarios. This includes: training a PAD system with live samples and spoof samples of one type of spoofing attack and then evaluating our model with a different - unseen - type of spoof samples; and training the PAD system with live samples only and test the model with live and various types of spoof samples.

The work presented here includes two main contributions: (i) a realistic estimation of the performance of binary classifiers in the presence of unseen spoofing attacks on iris recognition systems; (ii) the use of a single class approach. The single class classification is based on the use of decision models that rely only on the information from the real samples to detect the liveness. Such a model is depicted in Fig.2. Hence, for a fair comparison with the existing iris PAD techniques, three classification scenarios are evaluated.

*1) **One-Attack**:* This is the most commonly found PAD evaluation method in literature. In this scenario, it is assumed the availability of a training data set which has labelled instances for live and spoof classes for designing the classifier. Then, any unseen test data instance is compared against this trained model to determine which class it belongs to.

*2) **Unseen-Attack**:* Having in mind that in real-world solutions the system is not "aware" of the kind of spoofing attack that might be performed, in this scenario the known spoof samples are used to learn the model but complete knowledge is not assumed about the spoof samples. Therefore, the binary model is trained using samples obtained from several types of spoofing attacks. In the test phase a new type of samples is used to evaluate the model.

*3) **Single-Class**:* The single-class classification technique assumes that the training data has labeled instances for only the *live* class. The live samples are used to train the model but not the spoof ones. Since these models do not require labels for the spoof class, they are more widely applicable than the other approaches and do not overfit to the attacks in the training set.

III. Experimental Setup

A. Dataset

The experiments were performed using the GUC Visible Spectrum Iris Artefact (VSIA) Database [6] constructed for analyzing the effect of presentation attacks on visible iris recognition systems. Previous to this database there existed only one public artefact database of visible iris images, MobBIO*fake* [16]. The MobBIO*fake* dataset comprised attack data from only a single modality (print attack). Hence Unseen-Attack and Single-Class scenarios cannot be evaluated in that database. To the best of our knowledge there is no other database for iris liveness detection comprising samples from different types of spoofing attacks besides VSIA.

VSIA was constructed to generate iris artefact samples using both high definition electronic display screens and high quality printing. The five different presentations are: (i) Print Attack; (ii) Electronic Screen Attack Using iPad; (iii) Electronic Screen Attack Using Samsung Galaxy Tab; and Combining Print and Electronic Screen Attack - (iv) Using iPad and (v) Using Samsung Pad.

B. Feature Extraction Methods

Authors of the VISIA database compared various feature extraction techniques for "one-attack" scenario [6]. From the analysis carried out, the authors proposed to use Binarized Statistical Image Features (BSIF) to extract multiple features from the samples for different scales- both from the iris region and the peri-ocular region. A series of linear SVM classifiers were trained with each of these BSIF features. A voting scheme is used to combine the result of the classifiers. Such an approach yielded outstanding results for "One-Attack" scenario. But, it can be argued that this is a combination of multiple classification techniques and represents overfitting for this specific dataset. For fairness in comparison with other techniques, feature extraction for a single scale (7×7) **(BSIF7)** is incorporated in the experiments presented here.

Several other feature extaction methods were used to obtain a diversified set of discriminative features. The Weighted Local Binary Patterns **(wLBP)** method combines Local Binary Patterns (LBP) with a Scale Invariant Feature Transform (SIFT)

TABLE I. AVERAGE CLASSIFICATION ERROR RATE (ACER IN %) IN THE "ONE-ATTACK", "UNSEEN-ATTACK" AND "SINGLE-CLASS" SCENARIOS.

Features	Classification Scenario	ACER in % for Various Types of Attacks					Average ACER (in %)
		Attack 1	Attack 2	Attack 3	Attack 4	Attack 5	
wLBP	One-Attack	5.40	0.64	0.03	1.16	1.07	**1.66**
	Unseen-Attack	21.15	9.61	1.92	4.32	2.88	**7.98**
	Single-Class	31.57	26.28	13.30	19.07	20.67	22.15
WIris	One-Attack	12.46	0.00	0.00	0.00	1.44	2.78
	Unseen-attack	43.00	30.29	4.33	42.31	41.38	33.75
	Single-Class	28.69	2.56	2.40	12.34	15.38	12.28
EIris	One-Attack	12.46	0.00	0.00	0.46	1.44	2.87
	Unseen-attack	48.08	28.85	6.25	48.00	45.19	35.67
	Single-Class	34.62	2.08	0.96	5.45	11.70	**10.96**
LCP	One-Attack	10.72	0.10	0.00	2.77	6.67	4.05
	Unseen-attack	36.54	19.71	20.19	45.19	39.42	32.21
	Single-Class	34.61	24.04	10.74	14.10	21.31	20.96
BSIF7	One-Attack	19.69	0.00	0.00	0.56	2.87	4.36
	Unseen-attack	42.31	12.98	3.85	4.33	12.50	15.19
	Single-Class	35.74	21.31	20.35	29.49	39.26	29.23
Edg45	One-Attack	20.97	0.00	0.00	0.57	1.69	4.64
	Unseen-attack	35.10	8.65	43.27	18.75	13.94	23.94
	Single-Class	24.36	7.69	7.69	16.35	16.99	14.62
Edg90	One-Attack	24.15	0.15	0.15	1.79	2.62	5.77
	Unseen-attack	33.65	19.71	20.19	45.19	33.42	28.56
	Single-Class	24.67	24.04	10.74	14.10	21.31	20.96
LPQ	One-Attack	23.69	0.10	1.54	5.69	11.95	8.59
	Unseen-attack	32.21	14.90	20.67	17.79	27.40	22.60
	Single-Class	31.25	26.76	42.47	15.71	26.92	28.62

[17]. These features have shown high levels of performance in fingerprint liveness detection approaches [14].

The Eigen Iris feature extraction method **(EIris)** [18] and its variant **(WIris)** presented in [19] are learnt from the subspace of live iris images from the VSIA database. For both methods, a feature vector of 100 dimensions is used from the subspace analysis. These approaches have shown to perform well in face recognition systems [19].

The Edginess features are also used for the experiments in this work [19]. These features are extracted using one dimensional image processing for a specific angle from the horizontal axis. Features used in this work are extracted for 45 and 90 degrees, respectively **(Edg45** and **Edg90)**.

The Local Configuration Pattern **(LCP)** and Local Phase Quantization **(LPQ)** are extensively used in texture classification [20], [21]. LCP decomposes the information architecture of images into two levels and integrates both the microscopic features and local feature. LPQ is a textural method based on computing short-term fourier transform on a local image window. LCP and LPQ features are also analysed in this work.

C. Classifiers and parameter optimization

Support Vector Machines (SVM) and One-Class SVM (OCSVM) [22] are used as classifiers. In the "One-Attack" and "Unseen-Attack" scenarios, SVM with Linear and Radial Basis Function kernels are used. In the "Single-Class" scenario, OCSVM with a RBF kernel is used. SVM parameters are optimized using a grid-search by nested cross-validation.

In the PAD literature, SVM classifiers are preferred to other classifiers as they provide current state of- the-art results [6], [9], [10], [23]. In order to be consistent with the studies in literature and for a fair comparison, SVM based classifiers are used in this study. Nevertheless, other classifiers were tested in the preliminary studies and found SVM to be the optimum choice.

D. Performance Evaluation

The performance evaluation metrics used are the ones suggested by the standardization project, *ISO/IEC 30107-3 Presentation Attack Detection* [24]. The "Normal Presentation Classification Error Rate" (NPCER) is the proportion of normal/live presentations incorrectly classified as attack/spoof presentations and the "Attack Presentation Classification Error Rate" (APCER) is defined as the proportion of attack/spoof presentations incorrectly classified as normal/live presentations. The performance of the PAD algorithm is presented as the *Average Classification Error Rate (ACER)* which is given by the mean of the NPCER and the APCER error rates.

IV. EXPERIMENTAL RESULTS AND DISCUSSION

Average classification error rate (ACER) for the three classification scenarios - One-Attack, Unseen-Attack and Single-Class - are evaluated for five different types of presentation attacks present in VISIA database. Eight different types of feature extraction as discussed in Section III-B are evaluated and the results are presented in Table I.

From Table I, it can be observed that One-Attack classification scenario presents the best results. As discussed before, this is the most commonly used PAD technique and assumes the knowledge of all types of possible attacks. This previous knowledge of all possible attacks helped the classifier predicting the attack accurately. In the Unseen-Attack scenario, where a new type of attack is introduced apart from the ones used for training the classifier, an increase in ACER can be noted. This is expected as the classifier is optimized for decision making from the training data. One can also note that, these results are consistent with observation made by Bowyer and Doyle [15] in cosmetic contact lens detection.

When the classifier is learned with only images from the live iris (Single-Class approach), a decrease in ACER is noted for a majority of the feature extraction methods and attack types as compared to the Unseen-Attack scenario. This may be because the one-class classifier optimizes the decision boundary based on the live samples. No prior information of any attacks is used in this process and hence is robust when an unseen type of attack is presented at the system. This may be optimal for a real world iris recognition presentation attack detection technique. The results obtained by these last novel classification approaches clearly show that the traditional evaluation of models is overly optimistic about the different types of attacks that can be used.

Also, it can be noted that EIris features produced the lowest average ACER in Single-Class scenario. EIris features are learned from the subspace of live samples in VISIA database. This strengthen the argument that in a real world scenario, the PAD system should be designed mainly based on the information from the live subjects.

V. Conclusions and Future Work

This paper aims to question the approach of the majority of current presentation attack detection techniques for iris recognition systems, in particular by demonstrating that these tend to lead to an overly optimistic evaluation of system performance.

These traditional approaches assume prior knowledge of all presentation attacks types possible to train the classification system. Such an approach tends to lead to overfitting a specific database with limited range of attack samples often drawn from a single attack modality. It is shown in the Unseen-Attack scenario that when such a system is tested with a new type of presentation attack, the classification error rates increased significantly. A Single-Class classification scenario is recommended in this work, where both the features and classifier are learned from live samples only. Such a system does not require any prior knowledge of any possible attacks. This novel approach, even though resulting in a higher classification error rate when compared directly with the traditional one, leads to results which can be considered as more realistic in a real-world iris recognition system. It is recomended to evaluate the robustness of a liveness method to unseen spoof attacks by not assuming knowledge about the fake/spoof samples to be used by an intruder.

A future direction to follow is to design a presentation attack detection system mainly based on the information from the live subjects. Such a technique should not disregard completely the knowledge of existing presentation attacks that can be useful when similar attack methods are used.

References

[1] B. Toth, “Liveness detection: iris,” in *Encyclopedia of Biometrics*. Springer, 2009, pp. 931–938.

[2] J. Matey, O. Naroditsky, K. Hanna, R. Kolczynski, D. LoIacono, S. Mangru, M. Tinker, T. Zappia, and W. Zhao, “Iris on the move: Acquisition of images for iris recognition in less constrained environments,” *Proc. of the IEEE*, vol. 94, no. 11, pp. 1936–1947, Nov 2006.

[3] S. Thavalengal, P. Bigioi, and P. Corcoran, “Evaluation of combined visible/nir camera for iris authentication on smartphones,” in *IEEE Conf. on Comp. Vision and Patt. Recog. (CVPR) Workshops*, June 2015.

[4] M. D. Marsico, M. Nappi, and H. Proença, “Guest editorial introduction to the special executable issue on “Mobile Iris CHallenge Evaluation part I (MICHE I)”,” *Pattern Recognition Letters*, vol. 57, pp. 1 – 3, 2015.

[5] N. K. Ratha, J. H. Connell, and R. M. Bolle, “Enhancing security and privacy in biometrics-based authentication systems,” *IBM systems Journal*, vol. 40, no. 3, pp. 614–634, 2001.

[6] R. Raghavendra and C. Busch, “Robust scheme for iris presentation attack detection using multiscale binarized statistical image features,” *IEEE Transactions on Information Forensics and Security*, vol. 10, no. 4, pp. 703–715, April 2015.

[7] C. Sousedik and C. Busch, “Presentation attack detection methods for fingerprint recognition systems: A survey,” vol. 3, pp. 219 – 233, 2014.

[8] A. Czajka, “Pupil dynamics for iris liveness detection,” *Information Forensics and Security, IEEE Transactions on*, vol. 10, no. 4, pp. 726–735, 2015.

[9] A. F. Sequeira, H. P. Oliveira, J. C. Monteiro, J. P. Monteiro, and J. S. Cardoso, “MobILive 2014 - mobile iris liveness detection competition,” in *Proceedings of the Int. Joint Conference on Biometrics (IJCB)*, 2014.

[10] P. Gupta, S. Behera, M. Vatsa, and R. Singh, “On iris spoofing using print attack,” in *22nd International Conference on Pattern Recognition (ICPR), 2014*, Aug 2014, pp. 1681–1686.

[11] E. Marasco and C. Sansone, “On the robustness of fingerprint liveness detection algorithms against new materials used for spoofing.” in *BIOSIGNALS*, 2011, pp. 553–55 – 8.

[12] A. Rattani and A. Ross, “Automatic adaptation of fingerprint liveness detector to new spoof materials,” in *IEEE International Joint Conference on Biometrics (IJCB)*. IEEE, 29 September - 2 October 2014, pp. 1–8.

[13] A. Rattani, W. Scheirer, and A. Ross, “Open set fingerprint spoof detection across novel fabrication materials,” *IEEE Trans. on Inform. Forensics and Security,*, vol. 10, no. 11, pp. 2447–2460, Nov 2015.

[14] A. F. Sequeira and J. S. Cardoso, “Fingerprint liveness detection in the presence of capable intruders,” *Sensors*, 2015.

[15] K. Bowyer and J. Doyle, “Cosmetic contact lenses and iris recognition spoofing,” *Computer*, vol. 47, no. 5, pp. 96–98, May 2014.

[16] A. F. Sequeira, J. Murari, and J. S. Cardoso, “Iris liveness detection methods in mobile applications,” in *Proceedings of Int. Con. on Computer Vision Theory and Applications*, 2014, pp. 22 – 33.

[17] H. Zhang, Z. Sun, and T. Tan, “Contact lens detection based on weighted LBP,” in *20th International Conference on Pattern Recognition (ICPR)*, 23 - 26 August 2010, pp. 4279–4282.

[18] B.-R. Zheng, D.-Y. Ji, and Y.-H. Li, “Heterogeneous iris recognition using heterogeneous eigeniris and sparse representation,” in *2014 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, May 2014, pp. 3764–3768.

[19] T. Shejin and A. K. Sao, “Significance of dictionary for sparse coding based face recognition,” in *Proc. of the Int. Conf. of the Biometrics Special Interest Group (BIOSIG).*, Sept 2012, pp. 1–6.

[20] Y. Guo, G. Zhao, and M. Pietikäinen, “Texture classification using a linear configuration model based descriptor.” in *BMVC*. Citeseer, 2011, pp. 1–10.

[21] V. Ojansivu and J. Heikkil, “Blur insensitive texture classification using local phase quantization,” in *Image and Signal Processing, Lecture Notes in Computer Science*, vol. 5099, 2008, pp. 236 – 243.

[22] D. M. J. Tax and R. P. W. Duin, “Support vector domain description,” *Pattern Recognition Letters*, vol. 20, pp. 1191–1199, 1999.

[23] Y. Hu, K. Sirlantzis, and G. Howells, “Iris liveness detection using regional features,” *Pattern Recognition Letters*, pp. –, 2015.

[24] I. International Organization for Standardization, “ISO/IEC 5th WD 30107 : Information Technology - Biometrics - Presentation attack detection,” 2013.