IET Biometrics

Special Issue: Selected Papers from the International Workshop on Biometrics and Forensics (IWBF2017)

IET Journals
The Institution of Engineering and Technology

Non-reference image quality assessment and natural scene statistics to counter biometric sensor spoofing

ISSN 2047-4938
Received on 15th August 2017
Revised 28th November 2017
Accepted on 13th December 2017
E-First on 9th March 2018
doi: 10.1049/iet-bmt.2017.0146
www.ietdl.org

Dominik Söllinger[1], Pauline Trung[1], Andreas Uhl[1] ✉

[1]Department of Computer Sciences, University of Salzburg, Jakob-Haringer-Str. 2, 5020 Salzburg, Austria
✉ E-mail: uhl@cosy.sbg.ac.at

Abstract: Non-reference image quality measures (IQM) as well as their associated natural scene statistics (NSS) are used to distinguish real biometric data from fake data as used in presentation/sensor spoofing attacks. An experimental study shows that a support vector machine directly trained on NSS as used in blind/referenceless image spatial quality evaluator provides highly accurate classification of real versus fake iris, fingerprint, face, and fingervein data in generic manner. This contrasts to using the IQM directly, the accuracy of which turns out to be rather data set and parameter choice-dependent. While providing very low average classification error rate values for complete training data, generalisation to unseen attack types is difficult in open-set scenarios and obtained accuracy varies in almost unpredictable manner. This implies that for each given sensor/attack set-up, the ability of the introduced methods to detect unseen attacks needs to be assessed separately.

1 Introduction

We have observed a drastic increase in biometric authentication techniques being applied in various applications, ranging from border control to financial services. This is done to either complement or even replace classical authentication techniques based on tokens or passwords. Of course, this increased usage has also caused fraudulent attacks being mounted more often against biometric systems. Besides injecting fraudulent data into the communication inside a biometric system or attacking the template database, attacks against the proper functioning of the biometric sensor gain increasing importance. Such attacks are usually termed 'presentation' – or 'sensor-spoofing' – attacks and are conducted by presenting artefacts mimicking real biometrics traits to the biometric sensor to be deceived or by replaying earlier captured biometric sample data on some suited device, thus also attempting to deceive the sensor ('replay attack').

Counter-measures to this type of attacks have of course been considered already and are typically termed as 'anti-spoofing' or 'presentation-attack detection' measures [1]. In this context, very different approaches have been followed. The first type of anti-spoofing approach targets the *liveness* of the presented biometric traits in a passive or active manner and is thus termed as 'liveness-detection'. For example, pulse can be measured from facial video or hippus can be determined from temporal high-resolution iris video (in both cases, passive liveness detection is conducted). An example for active liveness detection is to determine reaction to illumination changes in pupil dilation during data acquisition for facial, periocular, or iris recognition systems. Passive liveness detection is efficiently able to prevent attacks conducted by e.g. gummy fingers or facial masks; however, it can be fooled by a replay attack as signs of liveness are also present in recaptured video. Active liveness detection on the other hand is able to withstand both types of attacks.

The second anti-spoofing approach directly focuses on the *replay* of previously recorded biometric sample data – as this attack involves the recapturing of previously recorded data by the biometric sensor the corresponding counter-measure is termed 'recapturing detection'. Techniques in this category include the detection of unnatural movement in video footage as indication of an attack, e.g. caused by hand motion when presenting a photo or display device to the sensor. Other approaches look into the interference between display refresh rate of the replaying device and temporal resolution of the video captured by the biometric sensor to detect an ongoing replay attack. Obviously, these methods are not able to detect attacks conducted with artefacts as they are directly and solely focused on the replay of the data.

The third type of anti-spoofing approach is more generic and uses *texture properties* of real biometric trait data acquired by the biometric sensor to discriminate from either recaptured data or data resulting from presenting some spoofing artefact to the sensor. Contrasting to liveness-based methods, which are specific to the target modality, and recapturing detection, which is limited and has to be focused to specific sensor/display type (including print-outs of course) combinations, texture-based methods usually employ generic texture descriptors together with subsequent machine learning techniques to discriminate real biometric data from spoofed variants. Of course, for this purpose, training data for classifier training is required. For example, a large variety of local image descriptors have been compared with respect to their ability to identify spoofed iris, fingerprint, and face data [2] and of course highly successful texture descriptors like local binary patterns have been extensively used for this purpose. However, it is often cumbersome to identify and/or design texture descriptors suited for a specific task in this context. Therefore, also generative techniques like deep learning employing convolutional neural networks have been successfully applied to discriminate real from spoofed biometric data [3, 4].

A very different way to identify spoofed data is to look into the quality of the imagery, assuming that the quality of the real biometric data is better or at least different from spoofed data. This of course can be seen as a specific type of texture-based discrimination approach. Related work considers two approaches in this context: first, the approach can be entirely agnostic of the considered modality by using general purpose image quality measures (IQM) [5, 6], and second, image quality metrics can be tailored to the biometric modality under investigation (see e.g. [7] which use face-specific data quality in order to recognise spoofing attacks against face recognition systems). The major contribution of this paper is to employ general purpose non-reference IQM (also termed 'blind' IQM) as well as the underlying natural scene statistics (NSS) in biometric spoofing attack/presentation attack detection and to assess their corresponding performance in different application settings. Complementing earlier results [5], we use (i) a different and larger set of non-reference IQM (six instead of two) and (ii) do not fuse the results with full-reference IQM values but focus on using one or several fused blind IQMs as generic spoofing detection technique.

Extending own prior work on using non-reference IQM for presentation attack detection [6, 8], (i) we add support vector machine (SVM) as a second classifier, also avoiding data-dependent parameter optimisation in its employment and thus achieving better result generalisability and present ISO/IEC 30107-3 compliant evaluation, (ii) we directly train blind/referenceless image spatial quality evaluator (BRISQUE) NSS on our data instead of using IQM output as classification input features, and (iii) we experimentally evaluate a specific type of open-set classification scenario, where our presentation attack detection schemes are confronted with real sample data of different sensors (i.e. looking into cross-sensor spoofing detection) and fake sample data of unseen subjects.

Section 2 introduces and explains the blind IQM as used in this paper. The databases specifically provided to test presentation attack detection techniques for iris, fingerprint, face, and fingervein recognition used in the present work are described in Section 3. Section 4 presents corresponding experimental anti-spoofing results in three distinct experimental set-ups, while Section 5 provides the conclusions of this paper.

2 Non-reference image quality metrics

Non-reference or blind IQM are easier in deployment when compared with full-reference or reduced-reference IQM as no information of the full-quality reference image is required for application. On the other hand, they are also harder to design as this lack of comparison data renders the design of these IQM much more difficult. There are different ways how to design blind IQM, depending on the necessity and type of training data used and the extent of generalisation potential of the admissible distortion types considered. Thus, depending on these design principles, we face some limitations. Among the techniques designed so far, we may distinguish opinion-aware (OA) IQM designs, where the IQM are trained on databases containing distorted imagery for which human annotations in terms of quality are available, and opinion-unaware (OU) IQM designs, which only rely on deviations from statistical regularities seen in natural images without the requirement of training on human annotated distortion databases. OA IQM are intrinsically limited as their assessment is limited to quality impairment resulting from distortion types they have been trained on. The examples for the first type, i.e. OA IQM, are distortion identification-based image verity and integrity evaluation (DIIVINE), blind image quality index (BIQI), and BRISQUE, while natural image quality evaluator (NIQE), blind image integrity notator (BLIINDS-II), and blind image quality assessment through anisotropy (BIQAA) are OU IQM.

Systematic comparisons of non-reference or blind IQM (NR IQM) as considered subsequently in spoofing detection have been published on traditional IQM tasks [9, 10]. Similarly, in non-trained [9] as well as in specifically trained manner [10], the correspondence to human vision is highly dependent on the target data set and on the nature of distortion present in the data. Thus, present studies did not identify a 'winner' among the techniques available concerning the correspondence to subjective human judgement and objective distortion strength.

2.1 OU NR image quality metrics

NIQE: The NIQE [11] is a spatial-domain IQM relying on an NSS model. The image is partitioned into patches for which sharpness is determined and only patches with sufficient sharpness are considered further. Those patches are pre-processed by local mean removal and divisive normalisation. From these data, for each patch, 36 NSS features are computed and these are fit to a multivariate generalised Gaussian (MVG) model. This MVG model is then compared to the 'natural' MVG model which is obtained by conducting the same procedure on natural images of good quality only. The extent of deviation from this model determines quality.

BLIINDS-II: The BLIINDS-II [12] computes NSS from a local discrete cosine transform (DCT) domain. After partitioning the image into patches, a local 2D DCT is computed on each of the blocks. Subsequently, the DCT domain in each block is partitioned into a low-frequency, mid-frequency, and high-frequency DCT subband, respectively. Furthermore, the DCT block is partitioned into three differently oriented subregions. Subsequently, an MVG fit is computed for each of the DCT subbands defined in this manner. From these parameters, the quality is derived in comparison to corresponding MVG parameters computed from high-quality imagery.

BIQAA: BIQAA [13] is the only NR IQM considered in this work which does not rely on NSS. In contrast, BIQAA measures the variance of the expected entropy of the image to be assessed in a set of predefined directions. Entropy is computed on a local basis by using a spatial-frequency distribution as an approximation for a pre-defined probability density function. For BIQAA, the generalised Renyi entropy and the normalised pseudo-Wigner distribution (PWD) are chosen in the used implementation. In this context, a pixel-by-pixel entropy value is computed enabling the generation of entropy histograms. The variance of the expected entropy is measured for different directions, and the differences are used to indicate anisotropy. Directional selectivity can be achieved by using an orientation-selective one-dimensional (1D) PWD implementation.

2.2 OA NR image quality metrics

BRISQUE: BRISQUE [14] operates in the spatial domain and uses virtually the same NSS as NIQE. The major difference to NIQE is the training on distorted images. For this purpose, similar kinds of distortions as present in the LIVE image quality database were introduced in each training image with varying strengths to create a set of the distorted images: JPEG 2000, JPEG, white noise, Gaussian blur, and fast fading channel errors. Subsequently, a mapping is learned from feature space to quality scores resulting in a measure of image quality. For that purpose, a SVM regressor is used.

DIIVINE: The DIIVINE [15] employs a two-stage framework consisting of distortion identification with subsequent distortion-specific quality determination. DIIVINE considers three common distortion types, i.e. JPEG compression, JPEG2000 compression, and blur.

In order to compute statistics from distorted images, the steerable pyramid decomposition is used. The steerable pyramid is an over-complete wavelet transform offering enhanced orientation selectivity when compared to using classical wavelet transform, as e.g. in BIQI.

BIQI: The BIQI [16] is based on a two-stage framework like DIIVINE as well and employs a classical wavelet transform over three scales using Daubechies 9/7 wavelet biorthogonal wavelet basis. The computed wavelet subband coefficients are used to compute NSS parameters (again an MVG fit is conducted): The first step is image distortion classification (which is based on a measure of how the NSS are modified and uses five distortion types: JPEG, JPEG2000, WN, Blur, and FF), the second step is quality assessment, using an algorithm specific to the distortion identified.

2.3 Natural scene statistics

IQM applied to images result in a single quality score in a certain range ([0, 100] in our set-up) for each IQM. Typically, these scores are obtained by applying machine learning techniques to map NSS to quality scores based on human judgement of distorted images or undistorted images only. Thus, the actual quality score delivered by IQM is neither directly related nor does it necessarily fit well to our application case for discriminating real from spoof biometric data. An alternative solution to this drawback is to avoid the deviation via quality scores but to train NSS directly on the 'real' and 'fake' labels of our data. Doing this, we also avoid the dimensionality reduction to a 1D quality score but retain the full NSS information for training.

3 Used spoofing/presentation attack databases

ATVS-FIr DB: The ATVS-Flr database consists of fake and real iris samples of both eyes of 50 subjects and complements the real data

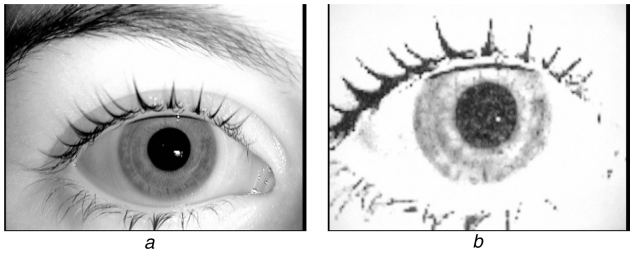

Fig. 1 *ATVS-FIr DB samples*
(a) Iris; right eye; real, ***(b)*** Iris; right eye; fake

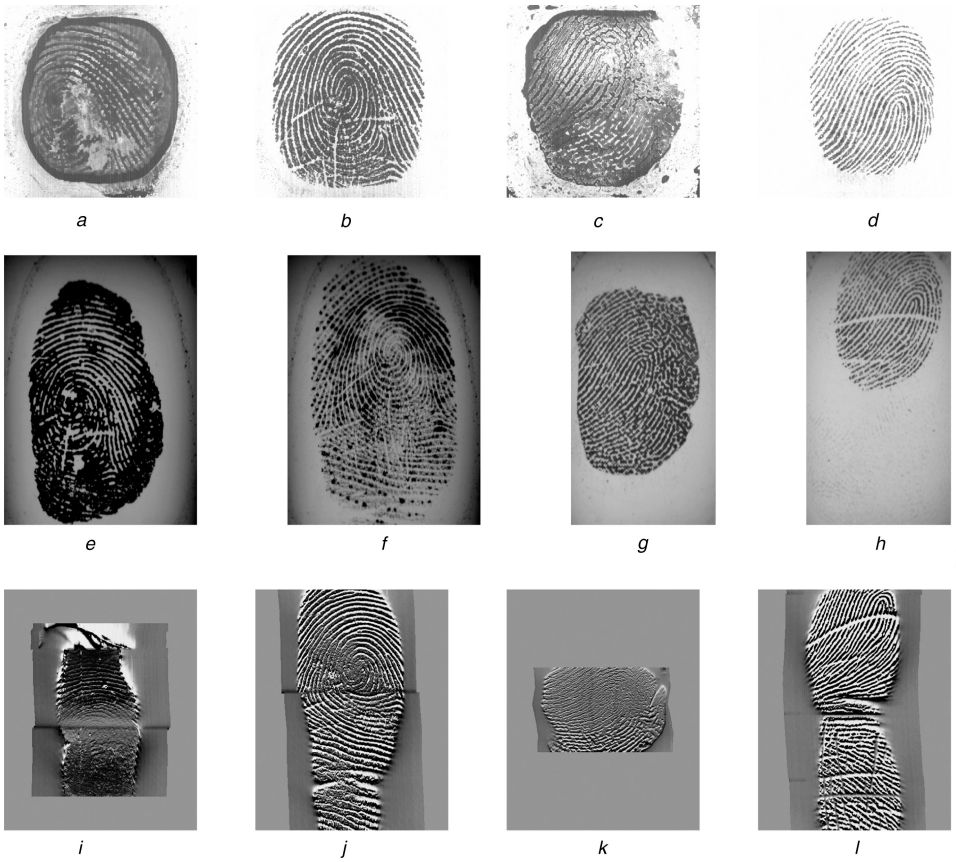

Fig. 2 *ATVS-FFp DB samples*
(a) WC; fake; capacitive, ***(b)*** WC; real; capacitive, ***(c)*** WOC; fake; capacitive, ***(d)*** WOC; real; capacitive, ***(e)*** WC; fake; optical, ***(f)*** WC; real; optical, ***(g)*** WOC; fake; optical, ***(h)*** WOC; real; optical, ***(i)*** WC; fake; thermal, ***(j)*** WC; real; thermal, ***(k)*** WOC; fake; thermal, ***(l)*** WOC; real; thermal

of the BioSecure data set [17]. Four samples of each iris were captured in two acquisition sessions with the LG Iris Access EOU3000. Thus, the database holds 800 real image samples (100 irises × 4 samples × 2 sessions). The fake samples were also acquired with the LG Iris Access EOU3000 from high-quality printed images of the original sample. As the structure is the same as for the real samples, the database comprises 800 fake image samples (100 irises × 4 samples × 2 sessions). Fig. 1 displays example images.

The data set has been used before in spoofing/presentation attack detection investigations, e.g. [2, 5, 18].

ATVS-FFp DB: The ATVS-FFp database consists of fake and real images taken from a human's index and middle finger of both hands. Those fingerprints can be divided into two categories: *with cooperation (WC)* and *without cooperation (WOC)*. 'WC' means that acquisition assumes the cooperation of the fingerprint owner, whereas images taken 'WOC' are latent fingerprints which had to be lifted from a surface.

Independent of the category, four samples of each finger were captured in one acquisition session with three different sensors:

- flat optical sensor Biometrika Fx2000 (512 dpi),
- sweeping thermal sensor by Yubee with Atmel's Fingerchip (500 dpi),
- flat capacitive sensor by Precise Biometrics model Precise 100 SC (500 dpi).

As a result, the database consists of 816 real/fake images (68 fingers × 4 samples × 3 sensors) samples taken WC and 768 real/fake images (64 fingers × 4 samples × 3 sensors) samples taken WOC. Fig. 2 displays example images from this data set.

The data set has been used before in spoofing/presentation attack detection investigations, e.g. in [19–21].

IDIAP replay-attack DB [22]: The replay-attack database for face spoofing consists of 1300 video clips of photo and video attack attempts to 50 clients under different lighting conditions. All videos were generated by either having a real client trying to access a laptop through its webcam or by displaying a photo/video to the webcam. Real as well as fake videos were taken under two different lighting conditions:

- Controlled: The office light was turned on, blinds are down, background is homogeneous.
- Adverse: Blinds up, more complex background, office lights are out.

Fig. 3 *IDIAP replay-attack DB samples*
(a) Adverse; real, ***(b)*** Adverse; fixed; fake highdef, ***(c)*** Adverse; hand; fake mobile, ***(d)*** Controlled; real, ***(e)*** Controlled; fixed; fake highdef, ***(f)*** Controlled; hand; fake mobile, ***(g)*** Adverse; fixed; fake print, ***(h)*** Adverse; hand; fake highdef, ***(i)*** Adverse; fixed; fake mobile

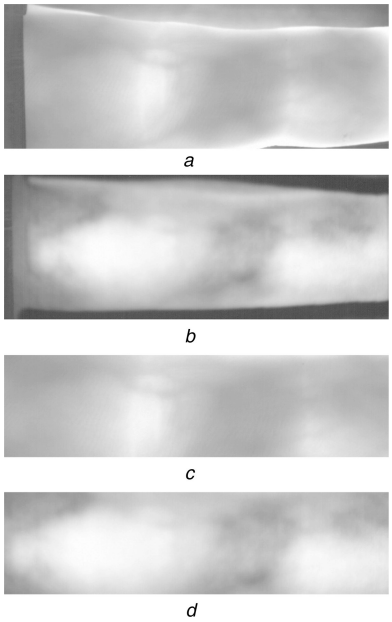

Fig. 4 *Finger vein DB samples*
(a) Full; real, ***(b)*** Full; fake, ***(c)*** Cropped; real, ***(d)*** Cropped; fake

To produce the attack, high-resolution videos were taken with a Canon PowerShot SX150 IS camera. The way to perform the attacks can be divided into two subsets: the first subset is composed of videos generated using a tripod to present the client biometry ('fixed'). For the second set, the attacker holds the device used for the attack with his/her own hands ('hand').

In total, 20 attack videos were registered for each client, 10 for each of the attacking modes just described:

- four times mobile attacks using an iPhone 3GS screen (with resolution 480 × 320 pixels),
- four times high-resolution screen attacks using an iPad (first generation, with a screen resolution of 1024 × 768 pixels),
- two times hard-copy print attacks (produced on a Triumph-Adler DCC 2520 colour laser printer) occupying the whole available printing surface on A4 paper.

As the algorithms used in our experiment are not compatible with videos, we extracted every *X*th frame from each video and used them as test data in our experiment. Fig. 3 displays example images used in experimentation.

The data set has been used before in spoofing/presentation attack detection investigations, e.g. in [2, 3, 5, 7, 22].

The spoofing-attack finger vein database [23]: This data set is provided by IDIAP Research Institute, consisting of 440 index finger vein images (both real authentications and spoofed ones (i.e. attack attempts)) corresponding to 110 subjects. Two different types of samples are available (as shown in Fig. 4): *full* (printed) images and *cropped* images where the resolution of the *full* images is 665 × 250 and that of the cropped images is 565 × 150 pixel, respectively.

This data set has been released in the context of the '1st Competition on Counter Measures to Finger Vein Spoofing Attacks' [23] and now it is the data basis for most research in finger vein sensor spoofing [24–26].

4 Experiments

4.1 Experimental set-up

For each image in the databases, quality scores were calculated with the IQM described in Section 2. We used the MATLAB implementations from the developers of BIQI, BLIINDS-2, NIQE, DIIVINE, BRISQUE (all available from http://live.ece.utexas.edu/research/quality/) and BIQAA (available at https://www.mathworks.com/matlabcentral/fileexchange/30800-blind-image-quality-assessment-through-anisotropy). In all cases, we used the default settings. We normalised the result data with the result that 0 represents a good quality and 100 the bad one which is already the default result in all cases except BIQAA. Originally,

the data of BIQAA is between 0 and 1. However, the values are so small that we had to define our own limits for the normalisation. A thorough analysis shows that our values are all between 0.00005 and 0.05; therefore, we used these figures as our limits. Moreover, we had to change the 'orientation' of the BIQAA quality scores to be conforming to our definition. Summarising, the following formula (1) was built:

$$x' = 100 - \frac{x - 0.05}{0.00005 - 0.05} \cdot 100 \quad (1)$$

4.1.1 Experiment 1: training sensor/setting identical to evaluation sensor/setting: In the first stage of experiment 1, we only consider the distribution of the quality scores. Our aim was to eventually find a threshold between the values of the real data and the fake ones for the various IQM.

Afterwards, in the second stage, we used the quality scores for a leave-one-subject-out cross-validation (training data is all data but the samples of the current subject to be classified, which is applied to each subject) to get an exact assertion about the classification possibility with NR IQM. To classify our data, we used *k*-nearest neighbours (kNN) as well as SVM classification. For kNN, our used *k* were 1, 3, 5, 7, and 9 (denoting the number of images with the closest feature vector considered) for this experiment and we exhaustively evaluate all combinations of IQM (i.e. resulting in different feature vector dimension and composition). Thus, we combined several quality scores of the different measures into one vector and used this for the kNN classification. The distance for the kNN-classification was the distance between the two vectors corresponding to the two images in question. The kNN results presented are the best ones, which means that we introduce a bias in the results here to see what is possible, but the best configuration in terms of IQM combination and *k* will be data dependent and will probably not generalise. For SVM, we use feature vectors consisting of all IQM scores (i.e. dimension 6) applying LIBSVM [27] with RBF kernel for training and thus, no bias by selecting certain IQM is introduced as all IQM are used. The grid-parameters (c, g) for the scalable vector regression were searched on a grid in logarithmic space. In order to conduct a fair evaluation in the used cross-validation, (c, g) are optimised within each training fold but then applied to the evaluation data.

The quantitative performance of the different techniques is measured according to the metric developed in ISO/IEC 30107-3 in terms of: (i) attack presentation classification error rate (APCER), which is defined as the proportion of an attack presentation incorrectly classified as normal (or real) presentation (false-negative spoof detection); (ii) normal presentation classification error rate (NPCER), which is defined as the proportion of a normal presentation incorrectly classified as an attack presentation (false-positive spoof detection). Finally, the performance of the overall technique is assessed in terms of average classification error rate (ACER):

$$\text{ACER} = \frac{\text{APCER} + \text{NPCER}}{2} \quad .$$

Of course, the lower the values of ACER (as well as APCER and NPCER), the better is the performance of the spoofing detection.

4.1.2 Experiment 2: training with BRISQUE NSS: In our experiment 2, we applied the BRISQUE NSS data and trained it on our labels. As first option, we applied kNN to the 36-dimensional BRISQUE NSS again using different values for *k*, presenting the best result achieved. As second option we applied SVM: The BRISQUE software does not only provide a pre-trained model delivering quality scores but also offers the option for training on different labels than quality scores using LIBSVM [27]. This is applied within the cross-validation evaluation.

4.1.3 Experiment 3: training sensor/setting different from evaluation sensor/setting: As correctly pointed out in [28], sensor spoof detection can of course not be considered a closed set problem. This means, that in a real-world scenario, the training data for a specific sensor will never be complete as in general we do not know which artefacts will be used by an attacker – thus the classifier should also work on unseen spoof types. This of course raises the question how to train a classifier based on such incomplete training data, a typical case of open set binary classification. The fact that the performance of a classifier will decrease when testing with samples unseen in training has been well studied in machine learning and pattern recognition, e.g. related to the 'over-fitting problem'. This generalisation problem of data-trained classifiers has been discussed also in the context of general image classification [29] and in biometrics (see e.g. in gender classification [30]). The general open-set recognition problem has recently been addressed [31–33] and the developed open-set classification techniques have been successfully applied to soft biometrics (mark, scar, and tattoo classification [34]), camera attribution and device linking [35], and fingerprint spoof detection [28]. In the latter work, emphasis is set to detect also attacks with unseen spoofing artefact fabrication material. Contrasting to that, one aspect of experiment 3 covers the issue of cross-sensor or inter-database spoofing detection. This means that unseen attacks involve samples acquired with different sensors than those the anti-spoofing system has been trained on, a topic that has gained increasing importance. Recent work has considered this scenario with various presentation attack detection methods for fingerprint [36–38], face [39], speaker [40], and iris [41, 42] recognition techniques, respectively.

In experiment 3, we investigated two different settings to simulate open-set scenarios: first, the ATVS-FFp database contains classical fingerprint imprints (WC) and latent fingerprints (WOC). So far, we have strictly separated those two sets as for both types real and fake versions are available. In order to simulate the open-set scenario, we trained the used classifier with classical imprints, while we evaluated on the latent fingerprint data. This can be seen as a special case of considering unseen fabrication material.

As a second setting, we used *real* sample data captured by *different sensors* and investigate how the spoof detection techniques trained on the sample data used before do react. In this setting, it is not entirely clear what to count as correct or incorrect decision (i.e. how to define APCER and NPCER): a (real) sample captured by a different sensor could be rated as 'real' as it corresponds to data captured from a real finger; on the other hand, it could be rated as 'fake' as it has been captured by a different sensor and might be the result of a successful injection attack. We follow the first consideration also due to the possibility to consider cross- and multi-sensor spoof detection techniques. Thus, a real sample captured by a different sensor should be rated correctly as being 'real', thus accumulating errors (samples rated as 'fake') in NPCER.

For iris samples, we used the *SDUMLA-HMT* data: This multi-modal data set was collected during the summer of 2010 at Shandong University, Jinan, China. One hundred and six subjects, including 61 males and 45 females with age between 17 and 31, participated in the data collecting process, in which five biometric traits – face, finger vein, gait, iris, and fingerprint, are collected for each subject [43]. SDUMLA-HMT is available at http://mla.sdu.edu.cn/sdumla-hmt.html. Every subject provided ten iris images, i.e. five images for each of the eyes.

For fingerprint samples, we employed samples of 49 individuals from the *CASIA-FingerprintV5* data set (http://biometrics.idealtest.org/dbDetailForUser.do?id=7) also used in [44, 45]. From these individuals, five samples per finger 1, 26, 7 are used.

Finally, for fingervein samples, two data sets are used. For *real* samples, we used the *UTFVP fingervein database* ([46], available at http://pythonhosted.org/bob.db.utfvp/), consisting of a total of 1440 images, taken from 60 subjects, 6 fingers per subject and 4 images per finger. For *fake* images, we used the *VERA spoofing fingervein database* (https://www.idiap.ch/dataset/vera-spoofingfingervein) consisting of 4 spoof images from 50 subjects. It has to be noted, that this data set has been acquired with the same sensor as the spoofing-attack finger vein database used in experiments 1 and 2 and consists only of fake data (only subjects

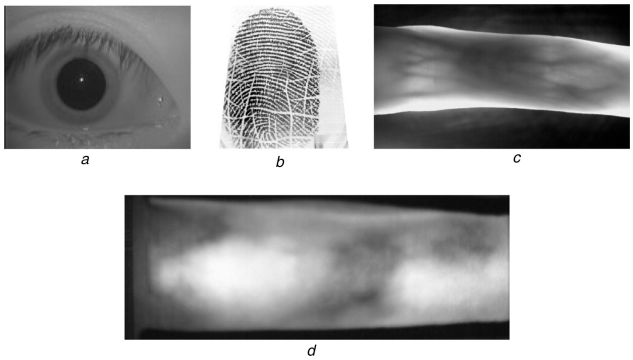

Fig. 5 *Examples for unseen data used in experiment 3*
(a) SDUMLA iris sample, ***(b)*** CASIA fingerprint sample, ***(c)*** UTFVP fingervein sample, ***(d)*** VERA fake fingervein sample

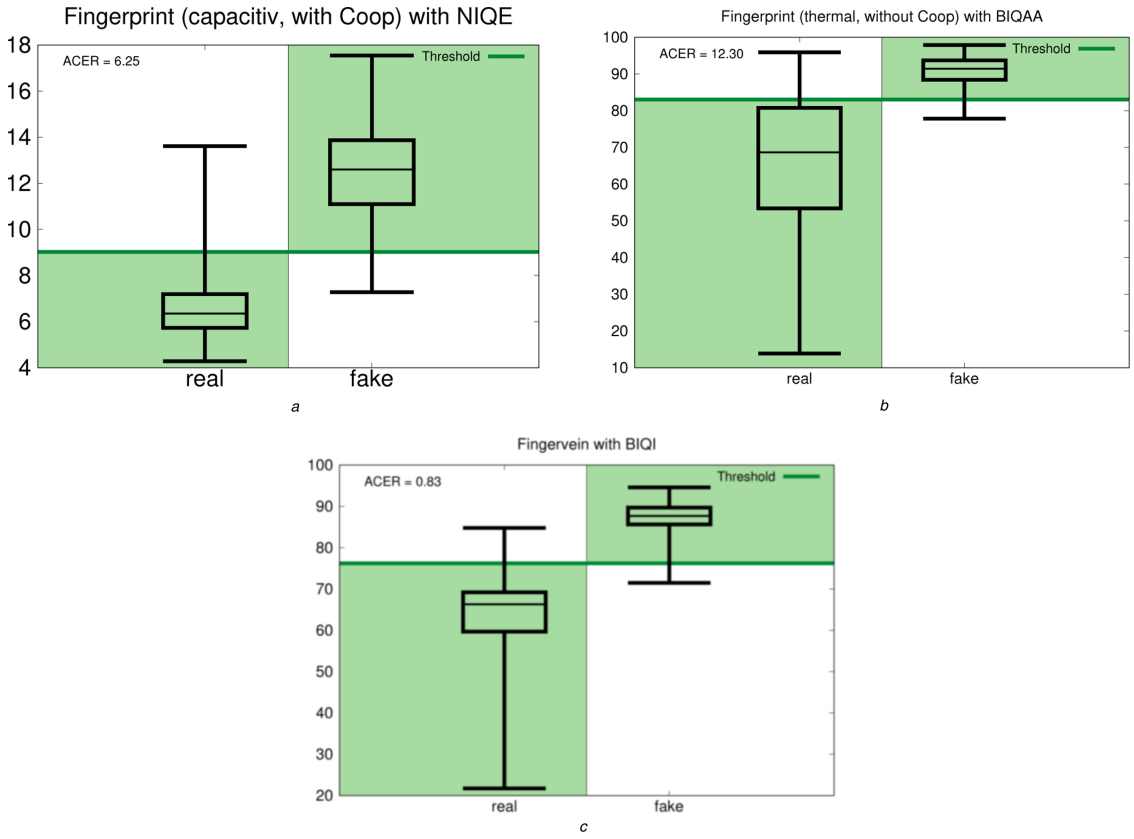

Fig. 6 *Quality score distribution (positive examples)*
(a) Fingerprint (capacitive, with coop) with NIQE, ***(b)*** Fingerprint (thermal, without coop) with BIQAA, ***(c)*** Fingervein (Full), with BIQI

differ between the two data sets). Thus, contrasting to the cases before, samples from this data set should be correctly classified as 'fake' and errors (i.e. samples rated as 'real') were counted in APCER. See Fig. 5 for an example of each data set.

Note, that in experiment 3, we have an intrinsic separation of training and evaluation data (contrasting to experiments 1 and 2). Therefore, we did not apply a leave-one-subject-out cross-validation but a direct classification of the evaluation samples based on the training data. In the second setting, involving the additional data sets containing real or fake data only, we only get NPCER or APCER results, thus ACER does not make sense and is omitted.

4.2 Experimental results

4.2.1 Experiment 1 – results: In Figs. 6 and 7, we display the distribution of single IQM values for real and fake data. For some cases, we notice a decent separation of the values almost allowing to specify a separation threshold. In the figures, we have depicted the threshold leading to the lowest ACER and have coloured the areas correctly classified in green. However, for most configurations, this simple strategy does not lead to useful results. In many cases (see e.g. Fig. 7), we could not recognise any separation between the distributions because they exhibited a similar mean and spread for real and the fake data. That was the reason for employing training-based classification techniques and fusion techniques.

In the case of kNN classification with only one IQM, we already obtain surprisingly good results [6, 8]. However, we were not able to identify a single IQM specifically well suited for the target task. In contrast, it seems that the different distortions present in the spoofed data are quite specific in terms of the nature and characteristic of the distortions, which is the only explanation of different IQM performing best on different data sets.

In fact, our results confirm the general results on IQM quality prediction performance [9, 10] in that it is highly data set and distortion-dependent which IQM provides the best results.

A further increase in classification accuracy (as computed by 100 – (APCER + NPCER)) is obtained by the combination of several IQM. Table 1 shows the best metric combinations in the case of kNN-classification for the considered databases from an exhaustive search. On average, we could improve our results by 7% compared to the single measure results [6, 8] and so most of the results are over 90%.

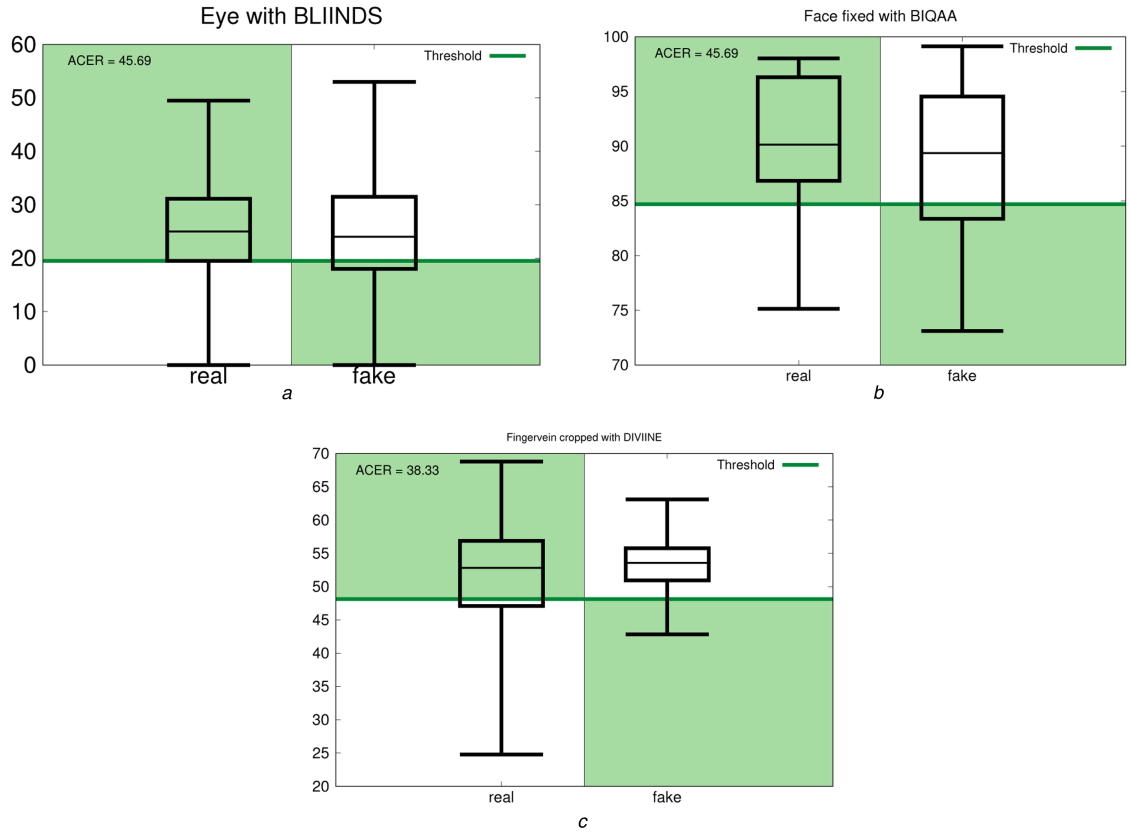

Fig. 7 *Quality score distribution (negative examples)*
(a) Eye with BLIINDS, ***(b)*** Face fixed with BIQAA, ***(c)*** Fingervein (cropped) with DIIVINE

Table 1 Best IQM combinations for kNN classification

Database	Combination	*k*	Accuracy, %
Iris	BIQI, BLIINDS, NIQE, DIIVINE, BRISQUE, BIQAA	9	85.81
fingerprint (optical, with coop)	BIQI, BLIINDS, DIIVINE, BIQAA	7	81.25
fingerprint (capacitive, with coop)	BIQI, BLIINDS, NIQE, DIIVINE, BRISQUE, BIQAA	3	96.69
fingerprint (thermal, with coop)	BIQI, BLIINDS, (NIQE), DIIVINE, BRISQUE	1	99.63
fingerprint (optical, without coop)	BIQI, BLIINDS, (NIQE), DIIVINE, BRISQUE	7	87.69
fingerprint (capacitive, without coop)	BLIINDS, NIQE, BIQAA	5	92.19
fingerprint (thermal, without coop)	BIQI, BLIINDS, NIQE, BRISQUE, BIQAA	1	98.44
face (hand)	BLIINDS, (NIQE), DIIVINE, BRISQUE, BIQAA	7	92.86
face (fixed)	BIQI, BLIINDS, NIQE, DIIVINE, BRISQUE, BIQAA	5	92.38
fingervein (full)	BIQI, DIIVINE, BRISQUE	1	99.79
fingervein (cropped)	BIQI, BLIINDS, NIQE, BRISQUE	9	85.63

From the latter table, we notice that there is a trend of getting best results when combining a larger number of IQM, confirming earlier results in this direction [5]. In order to look into this effect more thoroughly (and to clarify the role of the *k*-parameter in kNN classification), we have systematically investigated the results of the exhaustive classification scenarios in [6, 8]. We found that combining more metrics and choosing *k* large leads to better results on average, whereas the top results are achieved when using three to six metrics depending on the considered data set. For optimal values of *k*, we are not able to give a clear statement as *k* was also found to be 1 for three data sets in Table 1.

In Table 2, we display ISO/IEC 30107-3 compliant results comparing kNN and SVM classification. For correctly interpreting these results, it is important to consider that for kNN classification, we present the best result in terms of ACER achieved when considering all admissible values for *k* and all possible combinations of IQM. For the kNN case (left table half), we also provide the corresponding *k* value and the number of employed IQM in this best configuration. For SVM, we do not introduce any bias by including all six IQM score values into the feature vector.

The overall trend in terms of ACER is quite comparable for both kNN and SVM, as the databases exhibiting large and small ACER are identical for both techniques. For both kNN and SVM, there is no clear trend if APCER is usually larger as NPCER or vice versa. Also, there is no clear trend, which classification approach is better – SVM is superior for three data sets, while kNN is for six data sets. However, given that the kNN results come from a data-dependent parameter optimisation, SVM is strongly preferable as these results will generalise well as they are not at all fitted to the data. Overall, we face quite significant variations in terms of achieved ACER magnitude which implies that the methodology cannot be recommended as a general spoof detection approach but is restricted to suited data sets.

4.2.2 Experiment 2 – results: In Table 3, we show the results achieved when using BRISQUE NSS feature vectors instead of IQM ones, for both kNN as well as SVM classification. We observe identical behaviour with respect to the relation between APCER and NPCER as observed for IQM feature vectors (no clear trend which type of error is more frequent). When considering ACER, there is no clear improvement when changing from IQM to NSS feature vectors in the case of kNN classification.

Table 2 Comparing kNN and SVM classification using IQM features

	kNN-IQM				
Database	ACER	APCER	NPCER	*k*	# IQM
iris	7.09	5.69	8.5	9	6
fingerprint (optical, with coop)	9.38	12.87	5.88	7	4
fingerprint (capacitive, with coop)	1.65	1.29	2.02	3	6
fingerprint (thermal, with coop)	0.18	0.18	0.18	1	4
fingerprint (optical, without coop)	6.15	9.96	2.34	7	4
fingerprint (capacitive, without coop)	3.91	4.10	3.71	5	3
fingerprint (thermal, without coop)	0.78	0.98	0.59	1	5
face (hand)	3.57	4.29	2.86	7	4
face (fixed)	3.81	5.24	2.38	5	6
fingervein (full)	0.1	0	0.21	1	3
fingervein (cropped)	7.19	6.88	7.5	9	4

	SVM-IQM		
Database	ACER	APCER	NPCER
iris	2.22	4.06	0.38
fingerprint (optical, with coop)	13.14	13.26	13.05
fingerprint (capacitive, with coop)	1.19	0.74	1.65
fingerprint (thermal, with coop)	0.18	0.18	0.18
fingerprint (optical, without coop)	6.25	10.16	2.34
fingerprint (capacitive, without coop)	4.69	5.47	3.91
fingerprint (thermal, without coop)	0.88	0.59	1.17
face (hand)	4.52	5.71	3.33
face (fixed)	5.23	8.1	2.38
fingervein (full)	0.1	0	0.21
fingervein (cropped)	5.0	5.21	4.79

Table 3 Comparing kNN and SVM classification using NSS features

	kNN-NSS			
Database	ACER	APCER	NPCER	*k*
iris	0.88	1.06	0.69	1
fingerprint (optical, with coop)	7.90	7.72	8.09	5
fingerprint (capacitive, with coop)	0.83	0.92	0.74	1
fingerprint (thermal, with coop)	0.92	0.37	1.47	1
fingerprint (optical, without coop)	9.08	4.3	13.86	3
fingerprint (capacitive, without coop)	4.00	1.76	6.25	7
fingerprint (thermal, without coop)	0.78	0.59	0.98	7
face (hand)	5.95	5.71	6.19	5
face (fixed)	8.1	8.1	8.1	5
fingervein (full)	0.1	0.21	0	1
fingervein (cropped)	0.94	1.46	0.42	3

	SVM-NSS		
Database	ACER	APCER	NPCER
iris	0.06	0.06	0.06
fingerprint (optical, with coop)	0.64	0	1.29
fingerprint (capacitive, with coop)	0.28	0	0.55
fingerprint (thermal, with coop)	0.18	0.18	0.18
fingerprint (optical, without coop)	1.37	0.2	2.54
fingerprint (capacitive, without coop)	0.1	0.2	0
fingerprint (thermal, without coop)	0.29	0.2	0.39
face (hand)	2.14	0.48	3.81
face (fixed)	2.62	1.43	3.81
fingervein (full)	0	0	0
fingervein (cropped)	0.1	0.21	0

Table 4 Training with classical fingerprint data (WC), evaluation on latent fingerprints (WOC)

Database	kNN-IQM		
	ACER	APCER	NPCER
fingerprint (optical, without coop)	12.6	5.66	19.53
fingerprint (capacitive, without coop)	7.52	3.52	11.52
fingerprint (thermal, without coop)	5.66	1.56	9.77
database		kNN-NSS	
	ACER	APCER	NPCER
fingerprint (optical, without coop)	13.47	21.29	5.66
fingerprint (capacitive, without coop)	17.57	35.17	0
fingerprint (thermal, without coop)	2.54	4.49	0.59

Database	SVM-IQM		
	ACER	APCER	NPCER
fingerprint (optical, without coop)	18.65	2.15	35.16
fingerprint (capacitive, without coop)	9.77	0.59	18.95
fingerprint (thermal, without coop)	23.54	0.2	46.88
database		SVM-NSS	
	ACER	APCER	NPCER
fingerprint (optical, without coop)	25	50	0
fingerprint (capacitive, without coop)	23.34	46.68	0
fingerprint (thermal, without coop)	19.82	39.65	0

The situation is very different when considering the SVM results. NSS-based ACER values are clearly better for all but a single database (for which the values are identical) compared to IQM ones, partially considerably so. For example, ACER is reduced from 13.14 to 0.64 for the optical fingerprint data set (WC) and from 4.69/5 to 0.1 for both the capacitive fingerprint data set (WOC) and the full sized fingervein data set, respectively. Also, ACER values are superior for SVM compared to their kNN counterparts in the case of the NSS feature vectors. This is of particular interest, as the SVM results are expected to be highly generalisable due to the avoidance of data-specific bias. The significant superiority of SVM-NSS compared to kNN-NSS can probably be attributed to the significantly higher dimension of its feature vectors compared to SVM-IQM, for which SVN is much better able to exhibit its strengths as compared to kNN. As a consequence, we propose the employed SVM-NSS technique as a generic and rather accurate spoof detection methodology.

4.2.3 Experiment 3 – results: The set of last experiments is devoted to the open-set topic, i.e. looking into effects in case the type of evaluated samples are not part of the available training set. Table 4 shows the results when classifying real and fake latent fingerprints (denoted as WOC) when classification is based on classical fingerprint data (WC). We compare all four considered classification techniques in this table.

When comparing the obtained ACER results with the corresponding ones in Tables 2 and 3, we realise that in all four classification cases, ACER values are clearly worse in the 'open-set' scenario. Interestingly, worst ACER results are now exhibited for SVM-NSS, the approach clearly performing best in the 'closed-set' scenario. While this is surprising at first sight, it is not in fact. SVM-NSS is able to generate a very accurate model of the training data and thus performs quite well when working on seen spoof data. Contrasting, when confronted with unseen data very different from the training data, many errors do occur. Interestingly, not a single real sample is incorrectly classified as a fake one. However, almost every second fake sample is misclassified into a real one.

This is also a very different behaviour as seen with the closed-set scenario. In the open-set scenario, we observe significantly different magnitudes for APCER and NPCER, and the relation depends on the feature vector type. While for IQM-based feature vectors NPCER is clearly larger, the opposite is true for NSS-based ones.

Finally, in Table 5, we display results when confronting our spoof detection methodology with samples from unseen sensors or unseen subjects. As explained earlier, we only present NPCER or APCER values, as the employed data set only contain real or fake samples, but not both. Again, we compare the four classification methodologies considered so far. Additionally, kNN-IQM ∅ denotes the NPCER/APCER averaged over all results varying the number of used IQM exhaustively and taking the minimal value for $k = 1, 3, 5, 7, 9$. The aim is to show that average behaviour of the kNN behaviour may significantly deviate from the best results presented so far in the results. The results in the table clearly confirm this – average results are clearly worse as compared to the best ones as shown in the first column. In some cases, the difference is small (e.g. fingervein full), in other cases results change from perfect spoof detection to entirely useless results like for iris when changing from the best result to the average behaviour. This also implies that data dependency for kNN is rather high which leads to poor generalisation potential for this approach.

When looking at the results overall, we do hardly observe any general trends apart from the fact that results seem to strongly depend on the data sets considered and features/classification schemes employed. SVM-NSS, the classification scheme of choice for the closed-set scenario, performs perfectly for fingervein data, thus enabling cross-sensor spoof detection. On the other hand, for iris data, it does not work at all, classifying almost every real sample data as fake one while for fingerprint data every other real sample data is classified as fake. When looking at the actual pictorial data, it seems that fingervein data from different sensors is more similar than iris or fingerprint data from different sensors is (e.g. compare Figs. 4 and 5 for the fingervein case). NSS used with kNN classification exhibits the best overall results, with perfect classification for iris, three out of six fingerprint settings as well as for correctly detecting fake fingervein sample of unseen users but identical sensor. Applying SVM to IQM directly leads to consistent misclassifications in many cases, however, for two cases, the classification is almost perfect. The kNN results using IQM underpin the necessity of the k-parameter optimisation in case sensible results are expected. One might expect that corresponding fingerprint sensor types (i.e. optical versus optical) lead to lower error rates that different ones; however, the results do not reflect this behaviour. Overall, it is impossible to explain most effects in a sound manner, like the almost opposite behaviour for kNN and SVM on iris data for both feature vector types or the single outlier result of kNN-NSS for full fingervein data versus UTFVP. Also,

Table 5 NPCER/APCER (last line only) results when assessing sample data from different sensors/subjects

Database	kNN-IQM	kNN-IQM ∅	SVM-IQM	kNN-NSS	SVM-NSS
iris	0	68.72	89.11	0	100
fingerprint (optical, with coop)	9.59	35.18	100	60.41	50.55
fingerprint (capacitive, with coop)	22.35	81.28	100	99.38	50.20
fingerprint (thermal, with coop)	22.35	81.28	100	0	50
fingerprint (optical, without coop)	8.88	26.49	100	0	51.76
fingerprint (capacitive, without coop)	0.71	44.25	0.41	99.28	49.45
fingerprint (thermal, without coop)	0.71	44.25	100	0	51.76
fingervein (full versus UTFVP)	0	1.45	0	30.28	0
fingervein (full versus VERA)	80.5	99.69	100	0	0

the reasons for the entire failure of IQM feature vectors as opposed to NSS feature vectors for the full fingervein versus VERA data are hard to figure out.

5 Conclusion

We have found a high dependency on the actual data set/modality under investigation when trying to answer the question about the optimal settings when using non-reference IQM for biometric spoof detection. For some data sets, we obtain almost perfect separation of real and fake sample data, while for others, ACER values up to 10% can be observed.

The situation changes considerably, when directly training NSS features (in our experiments those used by BRISQUE) on our data, especially when using SVM classification. In this setting, worst ACER values are bound by 3.8%, with a majority of computed ACER values being significantly <1%, which makes this approach an interesting candidate for a generic spoof detection methodology.

In case the proposed spoof detection techniques are confronted with data from unseen sensors and/or subjects (modelling a more realistic open-set classification scenario incomplete training data), many results seem to be rather unpredictable. Thus, it seems to be advisable to apply recent open-set classification schemes to result in more stable and more generalisable results in case unseen data is to be expected.

6 Acknowledgments

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement no. 700259. Also, this work has been partially supported by the Austrian Science Fund, project no. 27776.

7 References

[1] Marcel, S., Nixon, M.S., Li, S.Z. (Eds.): '*Handbook of biometric anti-spoofing*' (Springer, London, 2014)

[2] Gragnaniello, D., Poggi, G., Sansone, C., *et al.*: 'An investigation of local descriptors for biometric spoofing detection', *IEEE Trans. Inf. Forensics Sec.*, 2015, **10**, (4), pp. 849–861

[3] Menotti, D., Chiachia, G., Pinto, A., *et al.*: 'Deep representations for iris, face, and fingerprint spoofing detection', *IEEE Trans. Inf. Forensics Sec.*, 2015, **10**, (4), pp. 864–879

[4] Raghavendra, R., Raja, K., Venkatesh, S., *et al.*: 'Transferable deep convolutional neural network features for fingervein presentation attack detection'. Proc. of the 5th Int. Workshop on Biometrics and Forensics (IWBF'17), Coventry, UK, 2017, pp. 1–6

[5] Galbally, J., Marcel, S., Fierrez, J.: 'Image quality assessment for fake biometric detection: application to iris, fingerprint, and face recognition', *IEEE Trans. Image Process.*, 2014, **23**, (2), pp. 710–724

[6] Bhogal, A.P.S, Söllinger, D., Trung, P., *et al.*: 'Non-reference image quality assessment for biometric presentation attack detection (best reviewed papers session)'. Proc. of the 5th Int. Workshop on Biometrics and Forensics (IWBF'17), Coventry, UK, 2017, pp. 1–6

[7] Wen, D., Han, H., Jain, A.K.: 'Face spoof detection with image distortion analysis', *IEEE Trans. Inf. Forensics Sec.*, 2015, **10**, (4), pp. 746–761

[8] Bhogal, A.P.S, Söllinger, D., Trung, P., *et al.*: 'Non-reference image quality assessment for fingervein presentation attack detection'. Proc. of 20th Scandinavian Conf. on Image Analysis (SCIA'17) (Springer Lecture Notes on Computer Science, **10269**), 2017, pp. 184–196

[9] Nouri, A., Charrier, C., Saadane, A., *et al.*: 'Statistical comparison of no-reference images quality assessment algorithms'. Proc. of the Colour and Visual Computing Symp. (CVCS'13), 2013

[10] Charrier, C., Saadane, A., Fernandez-Maloigne, C.: 'Comparison of no-reference image quality assessment machine learning-based algorithms on compressed images'. Image Quality and System Performance XII, volume 9396 of Proceedings of SPIE, 2015

[11] Mittal, A., Soundararajan, R., Bovik, A.C.: 'Making image quality assessment robust'. Proceedings of the 46th Asilomar Conf. on Signals, Systems and Computers (ASILOMAR), 2012

[12] Saad, M., Bovik, A.C., Charrier, C.: 'Blind image quality assessment: a natural scene statistics approach in the DCT domain', *IEEE Trans. Image Process.*, 2012, **21**, (8), pp. 3339–3352

[13] Gabarda, S., Cristobal, G.: 'Blind image quality assessment through anisotropy', *J. Opt. Soc. Am. A*, 2007, **24**, pp. 24–51

[14] Mittal, A., Moorthy, A.K., Bovik, A.C.: 'No-reference image quality assessment in the spatial domain', *IEEE Trans. Image Process.*, 2012, **21**, (12), pp. 4695–4708

[15] Moorthy, A.K., Bovik, A.C.: 'Blind image quality assessment: from natural scene statistics to perceptual quality', *IEEE Trans. Image Process.*, 2011, **20**, (12), pp. 3350–3364

[16] Moorthy, A.K., Bovik, A.C.: 'A two-step framework for constructing blind image quality indices', *IEEE Signal Process. Lett.*, 2010, **17**, (5), pp. 513–516

[17] Fierrez, J., Ortega-Garcia, J., Torre-Toledano, D., *et al.*: 'Biosec baseline corpus: a multimodal biometric database', *Pattern Recognit.*, 2007, **40**, (4), pp. 1389–1392

[18] Galbally, J., Ortiz-Lopez, J., Fierrez, J., *et al.*: 'Iris liveness detection based on quality related features'. Proc. of the IAPR/IEEE Int. Conf. on Biometrics (ICB'12), March 2012, pp. 271–276

[19] Galbally, J., Alonso-Fernandez, F., Fierrez, J., *et al.*: 'A high performance fingerprint liveness detection method based on quality related features', *Future Gener. Comput. Syst.*, 2012, **28**, pp. 311–321

[20] Bhardwaj, I., Londhe, N.D., Kopparau, S.K.: 'A spoof resistant multibiometric system based on the physiological and behavioral characteristics of fingerprint', *Pattern Recognit.*, 2017, **62**, pp. 214–224

[21] Lu, M., Chen, Z., Sheng, W.: 'Fingerprint liveness detection based on pore analysis'. Biometric Recognition – Proc. of the Chinese Conf. on Biometric Recognition (CCBR'15, Springer LNCS, 9428), 2015, pp. 233–240

[22] Chingovska, I., Anjos, A., Marcel, S.: 'On the effectiveness of local binary patterns in face anti-spoofing'. Proc. of the Int. Conf. of the Biometrics Special Interest Group (BIOSIG'16), September 2012

[23] Tome, P., Raghavendra, R., Busch, C., *et al.*: 'The 1st competition on counter measures to finger vein spoofing attacks'. 2015 Int. Conf. on Biometrics (ICB), May 2015, pp. 513–518

[24] Tirunagari, S., Poh, N., Bober, M., *et al.*: 'Windowed DMD as a microtexture descriptor for finger vein counter-spoofing in biometrics'. 2015 IEEE Int. Workshop on Information Forensics and Security (WIFS), November 2015, pp. 1–6

[25] Raghavendra, R., Busch, C.: 'Presentation attack detection algorithms for finger vein biometrics: a comprehensive study'. 2015 11th Int. Conf. on Signal-Image Technology Internet-Based Systems (SITIS), November 2015, pp. 628–632

[26] Kocher, D., Schwarz, S., Uhl, A.: 'Empirical evaluation of LBP-extension features for finger vein spoofing detection'. Proc. of the Int. Conf. of the Biometrics Special Interest Group (BIOSIG'16), Darmstadt, Germany, 2016, p. 8

[27] Chang, C.-C., Lin, C.-J.: 'LIBSVM: a library for support vector machines', *ACM Trans. Intell. Syst. Technol.*, 2011, **2**, (27), pp. 1–27. Available at http://www.csie.ntu.edu.tw/cjlin/libsvm

[28] Rattani, A., Scheirer, W.J., Ross, A.: 'Open set fingerprint spoof detection across novel fabrication materials', *IEEE Trans. Inf. Forensics Sec. (T-IFS)*, 2015, **10**, pp. 2447–2460

[29] Torralba, A., Efros, A.: 'Unbiased look at dataset bias'. Proc. of the IEEE Conf. on Computer Vision and Pattern Recognition (CVPR'11), 2011

[30] Guan, Y., Wei, X., Li, C.-T.: 'On the generalization power of face and gait in gender recognition', *Int. J. Digit. Crime Forensics*, 2014, **6**, (1), pp. 1–8

[31] Scheirer, W.J., Rocha, A., Sapkota, A., *et al.*: 'Towards open set recognition', *IEEE Trans. Pattern Anal. Mach. Intell. (T-PAMI)*, 2013, **35**, pp. 1757–1772

[32] Scheirer, W.J., Jain, L.P., Boult, T.E.: 'Probability models for open set recognition', *IEEE Trans. Pattern Anal. Mach. Intell. (T-PAMI)*, 2014, **36**, pp. 2317–2324

[33] Jain, L.P., Scheirer, W.J., Boult, T.E.: 'Multi-class open set recognition using probability of inclusion'. The European Conf. on Computer Vision (ECCV), September 2014

[34] Heflin, B., Scheirer, W.J., Boult, T.E.: 'Detecting and classifying scars, marks, and tattoos found in the wild'. The IEEE Int. Conf. on Biometrics: Theory, Applications and Systems (BTAS), September 2012

[35] Costa, F.O., Silva, E., Eckmann, M., *et al.*: 'Open set source camera attribution and device linking', *Pattern Recognit. Lett.*, 2014, **36**, pp. 92–101
[36] Akhtar, Z., Micheloni, C., Foresti, G.L.: 'Correlation based fingerprint liveness detection'. Proc. of the Int. Conf. on Biometrics (ICB'15), May 2015, pp. 305–310
[37] Marasco, E., Wild, P., Cukic, B.: 'Robust and interoperable fingerprint spoof detection via convolutional neural networks'. Proc. of the IEEE Symp. on Technologies for Homeland Security (HST'16), May 2016, pp. 1–6
[38] Chugh, T., Cao, K., Jain, A.K.: 'Fingerprint spoof detection using minutiae-based local patches'. Proc. of the Int. Joint Conf. on Biometrics (IJCB'17), 2017
[39] Patel, K, Han, H., Jain, A.K.: 'Cross-database face antispoofing with robust feature representation'. Proc. of the 11th Chinese Conf. on Biometric Recognition (CCBR'16), 2016, pp. 611–619
[40] Korshunov, P., Marcel, S.: 'Cross-database evaluation of audio-based spoofing detection systems'. Proc. of the Annual Conf. of the Int. Speech Communication Association, 2016, pp. 1705–1709
[41] Doyle, J.S., Bowyer, K.W., Flynn, P.J.: 'Variation in accuracy of textured contact lens detection based on sensor and lens pattern'. Proc. of the Int. Conf. on Biometrics: Theory, Applications and Systems (BTAS'13), September 2013, pp. 1–7
[42] Czajka, A., Bowyer, K.W., Krumdick, M., *et al.*: 'Recognition of image-orientation-based iris spoofing', *IEEE Trans. Inf. Forensics Sec.*, 2017, **12**, (9), pp. 2184–2196
[43] Yin, Y., Liu, L., Sun, X.: 'SDUMLA-HMT: A multimodal biometric database'. The 6th Chinese Conf. on Biometric Recognition (CCBR 2011), volume 7098 of Springer Lecture Notes on Computer Science, 2011, pp. 260–268
[44] Kirchgasser, S., Uhl, A.: 'Template ageing and quality analysis in time-span separated fingerprint data'. Proc. of the IEEE Int. Conf. on Identity, Security and Behavior Analysis (ISBA '17), New Delhi, India, 2017, pp. 1–8
[45] Kirchgasser, S., Uhl, A.: 'Template ageing in non-minutiae fingerprint recognition'. Proc. of the Int. Workshop on Biometrics and Forensics (IWBF '17), Coventry, UK, 2017, pp. 1–6
[46] Ton, B.T., Veldhuis, R.N.J.: 'A high quality finger vascular pattern dataset collected using a custom designed capturing device'. Int. Conf. on Biometrics, ICB 2013. IEEE, 2013