In Proceedings of the First Italian Conference on Cybersecurity (ITASEC17), Venice, Italy.

On the interoperability of capture devices in fingerprint presentation attacks detection

Luca Ghiani, Valerio Mura, Pierluigi Tuveri, and Gian Luca Marcialis

University of Cagliari
Department of Electrical and Electronic Engineering
{luca.ghiani,valerio.mura,pierluigi.tuveri,marcialis}@diee.unica.it

Abstract

A presentation attack consists in submitting to the fingerprint capture device an artificial replica of the finger of the targeted client. If the sensor is not equipped with an appropriate algorithm aimed to detect the fingerprint spoof, the system processes the obtained image as a one belonging to a real fingerprint. In order to face this problem, several presentation attacks detection (PAD) algorithms have been proposed so far. Current methods heavily rely on features extracted from a large data set of fake and real fingerprint images, and an appropriate classifier trained with such data to distinguish between live (real) and fake (spoof) fingerprint images. Building such data set requires a significant effort for fabricating samples of fake fingerprints, with the most effective materials used to circumvent the sensor. Interesting and promising results have been obtained, but they also suggest that the PAD is tailored on the particular sensor. Small and significant differences also occur when a novel version of the same sensor is released, and this may affect the PAD. Therefore, making a PAD interoperable is among the main current issues when considering fingerprints as the first level of protection and security of logical or physical resources. This paper is a first attempt to assess at which extent the sensor interoperability can be an issue for fingerprint PADs and to eventually propose a solution to this limitation. In particular, textural features will be under focus and a feature space transformation method based on the least square is proposed.

1 Introduction

Fingerprint capture devices suffer from the presentation attacks problem [4]. These attacks consist in submitting an artificial replica of the finger to the device. Fingerprint replica are fabricated by using, for example, silicon-based materials [12].

Fingerprint verification systems, without an appropriate detector of such spoofs, may process the fingerprint image as belonging to an authentic one. If the image quality and the replica are good enough, very similar to the original fingeprint, they can be circumvented [11].

This problem has been firstly pointed out in [12]. From then, many works appeared in literature to propose fingerprint presentation attacks detection (FPAD) algorithms, the most of them based on wavelet transform and filtering methods [16, 1, 14, 6, 5, 10]. At this regard, the international fingerprint FPAD competition (LivDet) in 2009, 2011, 2013 and 2015, tried to make the point about the effectiveness of such algorithms, by involving many academic and private institutions [7]. The main outcome of this competition has been more than ten data sets publicly available for research purposes, made up of images coming from a large set of sensors. Spoof images were derived by using different materials (from gelatine to liquid silicon)[1].

Although many promising results have been obtained, especially by features extracted with textural algorithms, there are significant issues still open. Among others, it is normal in fingerprint verification systems to substitute the capture device with a newest and better one,

[1] See http://livdet.diee.unica.it for further information on the competition and data sets.

Figure 1: Example of same fingerprint from Biometrika FX2000 (left), Italdata ET10 (right) scanner.

as a sort of upgrade. Unfortunately, this is not possible without updating templates as well[2]. In other words, there is not interoperability among fingerprint sensors, due to the different characteristic of image captured and, in some cases, on pre-processing steps operated before making available the image to the verification step [15, 2, 8].

In this paper, we answer the question: "is the interoperability also a problem for fingerprint presentation attacks detectors?". By an extensive experimental analysis on the data sets presented at LivDet 2011, we show that the problem exists (Section 2). On the other hand, we also show that the problem can be mitigated by an appropriate domain transformation algorithm, thus allowing to preserve the performance on the system during upgrade and downgrade (Sections 3-4). Conclusions are drawn in Section 5.

2 Interoperability experimental analysis

Each fingerprint scanner model is usually different from the others in both hardware and software. The main hardware difference is the sensor type: optical, solid state and ultrasound [9]. This peculiarity is linked to a different physical phenomenon, which codifies the valleys and ridges based on it. The scanners can be grouped by image characteristics such as DPI (dot per inch), scanning area, geometric accuracy. The DPI factor is a crucial point for matcher and liveness detector, it specifies the maximum resolution between two points. A high resolution scanner highlights details, such as pores, useful for FPAD. In the identification/verification task the DPI is very important for interdistance measurement of minutiae. A good example is the Bozorth matcher which works only at about 500DPI [18]. The scanning area defines the portion of captured fingertip, for example smartphone scanners are very small and do not acquire the entire fingerprint.

Software based differences are at API level, there can be many pre-processing steps in order to highlights ridges and valleys. Every step changes the dynamic of gray levels. The fingerprint matcher (comparator) based on the minutiae position, does not depend on these operations, as opposed to the performance of liveness detector that is based on a high frequency analysis. As a matter of fact the most important algorithms in FPAD like LBP [13], LPQ [6] and BSIF [5], work on a small local region of a gray scale image.

In Fig. 1 we are able to appreciate the differences of geometric distortion with two optical

[2]The template is a model stored in the data base and related to the registered subject, that is, authorized to pass the fingerprint verification test

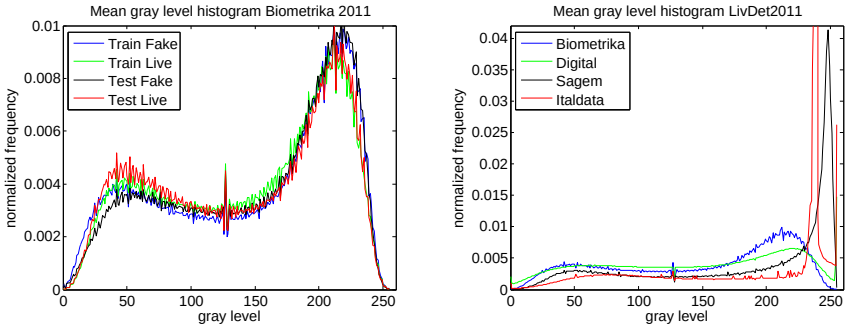

Figure 2: In the left side there is mean histogram of gray scale Biometrika for all subset. In the right side differences between mean histogram of LivDet2011 data set.

fingerprint capture devices, namely, Biometrika FX2000 and Italdata, adopted in the second edition of the International Competition on Figerprint Liveness Detection. The Biometrika capture device is more rounded than the Italdata one (both images belong to same fingerprint of the same person). The background is quite different, in the Italdata is almost white whereas in the Biometrika some shadows appear in the corners.

Another peculiarity is the gray level histogram. The example on the left side of Fig. 2 shows the image gray scale average histogram of Biometrika for the two subsets released for LivDet2011 and named training and test sets. In order to remove background effect, each fingerprint image is cropped, finding the center, in a ROI of 200x200 pixel. The trend of the four plots is quite similar. The main difference is in the highest gray levels between live and fake images. Moreover, on the right side of Fig. 2, differences among the four fingerprint scanners by plotting the average histogram of gray levels for all images (Biometrika, Digital Persona, Sagem, Italdata) are shown. As a consequence of those differences among sensors, we may expect that the feature vectors, calculated using for example textural algorithms, are different as well. The textural algorithms work very well in FPAD, but they can be used independently from the scanner. In other words, it may be expected no interoperability among scanners, because textural algorithms capture both small details of the fingerprint which are useful for our target and scanner-specific information.

In the following experiment we demonstrate this assertion using the four LivDet 2011 data sets [19], divided into a subset of 2000 live and 2000 fake samples.

Let us refer for example to two data sets coming from different sensors as A and B. We further subdivide these sets into train set $TrainA$ and test set $TestA$ such that $A = TrainA \cup TestA$ and $TrainB$ and $TestB$ such that $B = TrainB \cup TestB$. Let us then define the set $trainAB = TrainA \cup TrainB$ and the set as $testAB = testA \cup testB$. In this experiment, we train a classifier to discriminate among the two scanners images. In order to estimate the capability to discriminate the device, the training process is repeated with an increasing number of samples. The used classifier is a linear SVM [3], thus the sensor classes are separated by a decision hyperplane.

The used features are LBP [13], LPQ [6] and BSIF [5] histograms, that are the state of the art in fingerprint presentation attack detection.

Fig. 3 shows the mean accuracy rate for all the possible scanner pairs by increasing the number of training patterns. As expected, the performance grows up with the number of pattern, but even with a few samples the accuracy is more than 85%. Therefore, just a few patterns

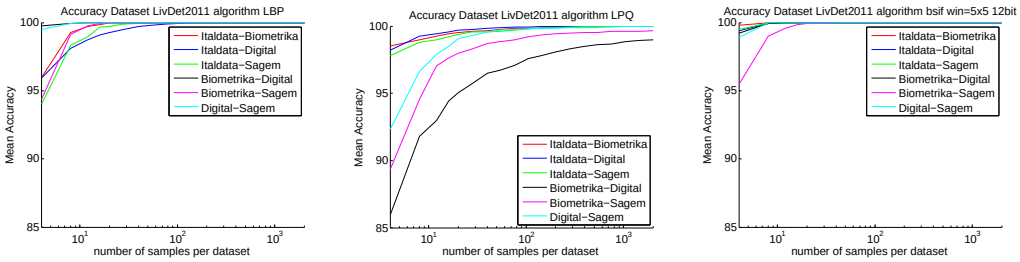

Figure 3: Accuracy of scanner recognition using LBP, LPQ and BSIF algorithms.

are sufficient to identify the capture device. This shows the existence of the interoperability problem.

	Biometrika test	Italdata test	Digital P. test	Sagem test
Biometrika train	88.85%	55.65%	64.35%	48.92%
Italdata train	51.20%	81.35%	65.65%	32.07%
Digital P. train	47.65%	50.55%	89.40%	50.59%
Sagem train	51.85%	48.20%	49.90%	91.65%

Table 1: LBP results: the accuracy are obtained training a classifier with the data set in the first column and testing on the data set in the first row.

	Biometrika test	Italdata test	Digital P. test	Sagem test
Biometrika train	85.20%	50.00%	57.75%	54.27%
Italdata train	55.90%	86.40%	59.00%	53.44%
Digital P. train	55.55%	50.00%	88.55%	54.66%
Sagem train	60.15%	50.00%	54.80%	92.78%

Table 2: LPQ results:the accuracy are obtained training a classifier with the data set in the first column and testing on the data set in the first row.

The second evidence is reported in Tables 1,2,3 where the liveness detection accuracy and cross-accuracy among fingerprint capture devices is reported. For example, the row 3 and the column 2 reports the liveness detection accuracy on features extracted from images coming from the Biometrika sensor submitted to the classifier trained on features extracted from images coming to the Italdata sensor. These values indicate that the features space of live and fake images is differently distributed for all scanners.

Based on these evidences, we can conclude that there is not interoperability among fingerprint capture devices even when liveness of fingerprint traits must be assessed. In other words, the liveness detector should be tailored on the specific sensor, by using an appropriate set of live and fake fingerprint images.

	Biometrika test	Italdata test	Digital P. test	Sagem test
Biometrika train	91.95%	50.00%	50.00%	52.50%
Italdata train	70.40%	86.15%	53.80%	59.72%
Digital P. train	50.00%	49.80%	95.85%	50.88%
Sagem train	53.05%	50.00%	50.00%	93.76%

Table 3: BISF results: the accuracy are obtained training a classifier with the data set in the first column and testing on the data set in the first row.

3 Feature space transformation: least square

In the previous sections we have shown, from a high level viewpoint, that it is not possible to substitute a sensor with a novel and better one without updating the whole liveness detection algorithm. In other words, the interoperability problem, already pointed out for fingerprint matching, also exists when upgrading a fingerprint liveness detection system. According to our results, this mainly affects the feature space distribution of live and fake images, especially if we adopt textural features, thus making impossible to keep the same classifier trained on features extracted from the "old" sensor. The practical consequence is that, if the new sensor must be used, a novel data set of live and fake fingerprints must be captured and a novel classifier must be trained. This can be quite expensive and difficult to obtain in a short time. Therefore, we propose here a method that can be adopted to mitigate the problem. "Mitigating" the problem means to allow users to adopt the new sensor whilst keeping the old classifier, by paying a drop of performance less significant that the one shown in the previous Section, until a novel data set is fully available.

To this aim, we used the least square method [17] that is adopted to minimize the error in the ill posed problem $A \times x = b$. In this problem A is an $n \times m$ matrix, x and b are two m and n elements vectors. Since we want to transform one $n \times m$ matrix into another, our problem can be divided in m steps:

$$A \times x_i = b_i \qquad i = 1, 2, ..., m \tag{1}$$

In each of these step we calculate the column vector x_i that, from the matrix A, generate the $i-th$ column vector of B.

Basically we select the features extracted from the datasets corresponding to two different sensors and calculate the matrix X that allows, with a simple product, to transform one feature set into the other by minimizing the squared error. In other words, we *adapt* the domain of the new sensor to that of the old one. We use an approximation algorithm instead of an interpolation algorithm because we want to avoid the overfitting of the new space on the old one.

Given the two matrices $trainA$ and $trainB$, according to the terms given in the previous Section, we calculate the matrix X such that:

$$trainB = trainA \times X \tag{2}$$

With this method we can turn every feature space A in any other feature space B by paying attention to the fact that we want to maintain the distinction between lives and fakes, that is, we want the live features of A moving on the live features of B and the fake features of A moving on the fake features of B. Thus, if each row of $trainA$ and $trainB$ contains a feature

vector, we have to be sure that for every row to each live (or fake) in $trainA$ correspond a live (or fake) in $trainB$. Clearly the two matrices must have the same number of rows (number of feature vectors) and columns (feature vector elements).

Let us suppose we have a large number of collected images with sensor B with which we trained a classifier and to have a new sensor A with which we want to build another dataset. If the number of images acquired with the new sensor is limited, can we exploit the other classifier moving the feature space of A in that of B? Can we just use live fingerprints or is it required to also collect fake samples? How many feature vectors do we need to get satisfying results? In order to answer these questions we performed a number of experiments by analyzing the LivDet 2011 datasets [19]. Features were extracted with three different textural algorithms: LBP[13], LPQ[6] and BSIF[5]. For sake of space we will only show the results obtained with the LBP algorithm. Results on other algorithms were similar and they can also be requested to the authors.

3.1 Least square transformation

In our first experiment we used all possible pairs of LivDet2011 datasets. For each couple of datasets A and B, given the $trainA$ and $trainB$ matrices, we calculate the transformation matrix X (see Eq. 2) that allows us to move a vector from the A feature space to the B one. With the X and the $testA$ matrices we are easily able to compute a $pseudoTestB$ matrix moving the $testA$ feature vectors in the B feature space:

$$pseudoTestB = testA \times X \tag{3}$$

In Table 4 we present the obtained results in terms of accuracy. The values in the diagonal, since A and B represent the same dataset, are obtained with the classical procedure and without any transformation: we just trained a classifier with $trainB$ and we used it to classify the feature vectors in $testB$. Conversely, for each off-diagonal value, we still trained a classifier with $trainB$ (rows), but then we calculated the matrix X by Eq. 2 and the $pseudoTestB$ by Eq. 3. Finally, we classified the feature vectors in $pseudoTestB$ (columns).

	Biometrika pseudo-test	Italdata pseudo-test	Digital P. pseudo-test	Sagem pseudo-test
Biometrika train	88.85%	80.10%	84.15%	86.25%
Italdata train	86.50%	81.35%	85.60%	88.80%
Digital train	86.55%	79.65%	89.40%	85.07%
Sagem train	83.90%	77.15%	89.35%	91.65%

Table 4: LBP results: the accuracies are obtained by training a classifier with the datasets in the first column and testing on the datasets in the first row.

The presented results clearly show that the accuracies obtained by transforming $testA$ in a $pseudoTestB$ are comparable (but almost always lower) with those obtained by simply training the original dataset $trainB$. In other words, the effect of the passage from a sensor to another one is mitigated or reduced.

3.2 Transformation using only live samples

The results in Section 3.1 are obtained using both lives and fakes feature vectors to calculate the transformation matrix. Since the acquisition of new fakes is not as easy and fast as collecting

live samples, which is not a trivial problem as well, we repeated the experiments using the live samples to calculate the transformation matrix.

	Biometrika pseudo-test	Italdata pseudo-test	Digital P. pseudo-test	Sagem pseudo-test
Biometrika train	88.85%	57.75%	50.00%	49.41%
Italdata train	72.30%	81.35%	50.00%	50.88%
Digital train	49.95%	50.00%	89.40%	49.17%
Sagem train	50.00%	50.00%	50.00%	91.65%

Table 5: LBP results: the accuracies are obtained training a classifier with the datasets in the first column and testing on the datasets in the first row were the pseudo-tests are calculated with using only live samples.

Unfortunately, as shown in Table 5 the results are much worse, but this is also explainable since we have willingly reduced the information available for the transformation, by avoiding the use of fake samples. Whilst this is an indirect evidence that textural algorithms are able to extract live and fake fingerprint characteristics, this also show that it is not possible to avoid this information when designing the domain adaptation function. As a matter of fact, the majority of fake fingerprints were classified as live ones in our experiments.

3.3 Number of feature vectors

In the previous subsections the transformation matrices have been calculated using all the feature vectors (lives and fakes in Section 3.1 and just lives in Section 3.2) extracted from the images in the train parts of the LivDet 2011 datasets [19] (approximately 2000, 1000 lives and 1000 fakes). In order to simulate a limited number of acquisition, we replicate the experiments using just a randomly selected subset of those feature vectors (with the same number of lives and fakes in both subset). Given the two matrices $trainA$ and $trainB$ we extract the two subset $subSetA$ and $subSetB$ and calculate the new matrix X such that:

$$subSetA = subSetB \times X \tag{4}$$

In the experiments presented in Tables 6 - 9 the transformation matrices were calculated with subsets containing the 20%, 40%, 60% and 80% of the feature vectors of the original train datasets.

	Biometrika pseudo-test	Italdata pseudo-test	Digital P. pseudo-test	Sagem pseudo-test
20%	86.51%	73.20%	85.43%	82.28%
40%	87.25%	76.48%	87.82%	86.83%
60%	87.69%	76.13%	89.01%	87.89%
80%	88.64%	77.71%	89.92%	87.91%

Table 6: LBP results: the accuracies are obtained training a classifier with the Biometrika dataset and testing on the datasets in the first row obtained using a transformation matrix calculated with the train percentages in the first column.

Results show, as expected, an increasing trend of the accuracies since the bigger is the train percentage, the higher is usually the recognition rate. Anyway, in most cases, a 40% of the

	Biometrika pseudo-test	Italdata pseudo-test	Digital P. pseudo-test	Sagem pseudo-test
20%	79.65%	75.97%	81.74%	79.41%
40%	82.72%	79.27%	87.43%	84.79%
60%	83.40%	81.40%	86.98%	86.82%
80%	84.22%	80.71%	88.57%	87.52%

Table 7: LBP results: the accuracies are obtained training a classifier with the Italdata dataset and testing on the datasets in the first row obtained using a transformation matrix calculated with the train percentages in the first column.

	Biometrika pseudo-test	Italdata pseudo-test	Digital P. pseudo-test	Sagem pseudo-test
20%	72.27%	69.41%	86.48%	82.76%
40%	78.35%	72.75%	88.41%	85.41%
60%	82.03%	76.79%	88.72%	86.42%
80%	83.16%	75.18%	89.14%	86.97%

Table 8: LBP results: the accuracies are obtained training a classifier with the Digital Persona dataset and testing on the datasets in the first row obtained using a transformation matrix calculated with the train percentages in the first column.

train seems to be sufficient to obtain satisfying results. This means that it is possible to keep the old classifier without an appreciable loss of performance and using at the same time the new sensor, until a good number of live and fake samples has been collected.

4 Conclusions

In this paper, we analyzed the level of interoperability between sensors in the field of fingerprint presentation attacks (liveness) detection. We showed that the problem exists and that it is difficult to overcome since the sensor characteristics heavily affect the image characteristics and the related feature space when using textural algorithms. These features contain so much "sensor-specific" information that the related images can be localized in different regions of the features space itself, almost without overlapping.

Starting from these observations we proposed a method to strongly reduce these sensor-specific features distributions. This method is based on the least square algorithm, pros and cons were also discussed in the paper. Although the pointed out limitations, it allowed achieving a considerable level of interoperability among fingerprint capture devices.

In the future we will focus our efforts on the improvement of the performances but also on the reduction of the number of feature vectors required to calculate the transformation matrix. In particular, calculating the domain transformation equation without the need to use fake samples would be of great help to design an effective algorithm for capture devices interoperability in fingerprint liveness detection.

	Biometrika pseudo-test	Italdata pseudo-test	Digital P. pseudo-test	Sagem pseudo-test
20%	80.64%	72.67%	86.53%	89.12%
40%	80.83%	75.38%	88.72%	91.06%
60%	82.00%	75.22%	89.24%	91.47%
80%	80.34%	77.11%	89.49%	91.51%

Table 9: LBP results: the accuracies are obtained training a classifier with the Sagem dataset and testing on the datasets in the first row obtained using a transformation matrix calculated with the train percentages in the first column.

References

[1] A. Abhyankar and S. Schuckers. A wavelet-based approach to detecting liveness in fingerprint scanners. In *Defense and Security*, pages 278–286. International Society for Optics and Photonics, 2004.

[2] Fernando Alonso-Fernandez, Raymond NJ Veldhuis, Asker M Bazen, Julian Fiérrez-Aguilar, and Javier Ortega-Garcia. Sensor interoperability and fusion in fingerprint verification: a case study using minutiae-and ridge-based matchers. In *2006 9th International Conference on Control, Automation, Robotics and Vision*, pages 1–6. IEEE, 2006.

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

[4] Javier Galbally, Julian Fierrez, Javier Ortega-Garcia, and Raffaele Cappelli. Fingerprint anti-spoofing in biometric systems. In *Handbook of Biometric Anti-Spoofing*, pages 35–64. Springer, 2014.

[5] L. Ghiani, A. Hadid, G. L. Marcialis, and F. Roli. Fingerprint liveness detection using binarized statistical image features. In *Biometrics: Theory, Applications and Systems (BTAS), 2013 IEEE Sixth International Conference on*, pages 1–6, Sept 2013.

[6] L. Ghiani, G. L. Marcialis, and F. Roli. Fingerprint liveness detection by local phase quantization. In *Pattern Recognition (ICPR), 2012 21st International Conference on*, pages 537–540, Nov 2012.

[7] Luca Ghiani, David A Yambay, Valerio Mura, Gian Luca Marcialis, Fabio Roli, and Stephanie A Schuckers. Review of the fingerprint liveness detection (livdet) competition series: 2009 to 2015. *Image and Vision Computing*, 2016.

[8] Luca Lugini, Emanuela Marasco, Bojan Cukic, and Ilir Gashi. Interoperability in fingerprint recognition: A large-scale empirical study. In *Dependable Systems and Networks Workshop (DSN-W), 2013 43rd Annual IEEE/IFIP Conference on*, pages 1–6. IEEE, 2013.

[9] Davide Maltoni, Dario Maio, Anil K. Jain, and Salil Prabhakar. *Handbook of Fingerprint Recognition*. Springer Publishing Company, Incorporated, 2nd edition, 2009.

[10] Emanuela Marasco and Arun Ross. A survey on antispoofing schemes for fingerprint recognition systems. *ACM Computing Surveys (CSUR)*, 47(2):28, 2014.

[11] T. Matsumoto. Gummy and conductive silicone rubber fingers. In *Advances in Cryptology - ASIACRYPT 2002, 8th International Conference on the Theory and Application of Cryptology and Information Security, Queenstown, New Zealand, December 1-5, 2002, Proceedings*, pages 574–576, 2002.

[12] T. Matsumoto, H. Matsumoto, K. Yamada, and S. Hoshino. Impact of artificial gummy fingers on fingerprint systems. *Datenschutz und Datensicherheit*, 26(8), 2002.

[13] S. B. Nikam and S. Agarwal. Texture and wavelet-based spoof fingerprint detection for fingerprint biometric systems. In *2008 First International Conference on Emerging Trends in Engineering and Technology*, pages 675–680, July 2008.

[14] Shankar Bhausaheb Nikam and Suneeta Agarwal. Local binary pattern and wavelet-based spoof fingerprint detection. *International Journal of Biometrics*, 1(2):141–159, 2008.

[15] Arun Ross and Anil Jain. *Biometric Sensor Interoperability: A Case Study in Fingerprints*, pages 134–145. Springer Berlin Heidelberg, Berlin, Heidelberg, 2004.

[16] S. Schuckers. Spoofing and anti-spoofing measures. *Inf. Sec. Techn. Report*, 7(4):56–62, 2002.

[17] L. Ridgway Scott. *Numerical Analysis*. Princeton University Press, Princeton, NJ, USA, 2011.

[18] Craig I. Watson, Michael D. Garris, Elham Tabassi, Charles L. Wilson, R. Michael Mccabe, Stanley Janet, and Kenneth Ko. User's guide to nist biometric image software (nbis).

[19] D. Yambay, L. Ghiani, P. Denti, G. L. Marcialis, F. Roli, and S. Schuckers. Livdet 2011 - fingerprint liveness detection competition 2011. In *2012 5th IAPR International Conference on Biometrics (ICB)*, pages 208–215, March 2012.