SCISPACE
formerly Typeset

Open access • Proceedings Article • DOI:10.1109/ICB.2013.6612981

Can face anti-spoofing countermeasures work in a real world scenario?

— Source link

Tiago de Freitas Pereira, André Anjos, José Mario De Martino, Sébastien Marcel

Institutions: State University of Campinas, Idiap Research Institute

Published on: 04 Jun 2013 - International Conference on Biometrics

Topics: Spoofing attack, Face Recognition Grand Challenge, Biometrics and Replay attack

Related papers:

- On the effectiveness of local binary patterns in face anti-spoofing
- A face antispoofing database with diverse attacks
- Face Spoof Detection With Image Distortion Analysis
- Face spoofing detection from single images using micro-texture analysis
- Face liveness detection with component dependent descriptor

Share this paper:

View more about this paper here: https://typeset.io/papers/can-face-anti-spoofing-countermeasures-work-in-a-real-world-4wsg2u0fuc

Can face anti-spoofing countermeasures work in a real world scenario?

Tiago de Freitas Pereira[1 2], André Anjos[3], José Mario De Martino[1], Sébastien Marcel[3]
[1]School of Electrical and Computer Engineering - University of Campinas (UNICAMP),
[2]CPqD Telecom & IT Solutions,
[3]IDIAP Research Institute
tiagofrepereira@gmail.com, andre.anjos@idiap.ch, martino@fee.unicamp.br, marcel@idiap.ch

Abstract

User authentication is an important step to protect information and in this field face biometrics is advantageous. Face biometrics is natural, easy to use and less human-invasive. Unfortunately, recent work has revealed that face biometrics is vulnerable to spoofing attacks using low-tech equipments. This article assesses how well existing face anti-spoofing countermeasures can work in a more realistic condition. Experiments carried out with two freely available video databases (Replay Attack Database and CASIA Face Anti-Spoofing Database) show low generalization and possible database bias in the evaluated countermeasures. To generalize and deal with the diversity of attacks in a real world scenario we introduce two strategies that show promising results.

1. Introduction

Because of its natural and non-intrusive interaction, identity verification and recognition using facial information are among the most active and challenging areas in computer vision research. Despite the significant progress of face recognition technology in the recent decades, wide range of viewpoints, ageing of subjects and complex outdoor lighting are still research challenges. Advances in the area were extensively reported in [1, 2]. However, checking if the face presented to a camera is indeed a face from a real person and not an attempt to deceive (spoof) the system has mostly been overlooked.

A spoofing attack consists of the use of forged biometric traits to gain illegitimate access to secured resources protected by a biometric authentication system. Recently, the media has documented some situations of attacks in deployed face recognition systems. Using simple photographs, a research group from University of Hanoi showed how easy is to spoof the face authentication systems deployed in Lenovo, Asus and Toshiba Laptops [3]. Since the release *Ice Cream Sandwich*, the Android OS come with a built-in face authentication system to unlock the mobile phone. Since then, it has been extensively demonstrated around the web how easy it is to spoof this face recognition system[1]. As a consequence, an eye blinking detection has been introduced in the most recent version of the Android OS.

The lack of protection against biometric spoofing attacks is not exclusive to face biometrics. The findings in [4, 5] indicate that fingerprint authentication systems suffer from a similar weakness. The same shortcomings on iris recognition systems have been diagnosed [6, 7]. The literature review for spoofing in face recognition systems is presented in Section 2.

While it is possible to spoof a face authentication system using make-up, plastic surgery or forged masks; photographs and videos are probably the most common threats. Moreover, due to the increasing popularity of social network websites (facebook, youtube, flickr, instagram and others), a great deal of multimedia content, specially videos and photographs, is available on the web that can be used to spoof a face authentication system. To mitigate vulnerabilities, effective countermeasures against face spoofing must to be deployed.

It was not until very recently that the problem of spoofing attacks against face biometric systems gained attention of the research community. This can be attested by the gradually increasing number of publicly available databases [8, 9, 10] and the recently organized IJCB 2011 competition on countermeasures to 2D facial spoofing attacks [11]. This was the first competition conducted for studying best practices for non-intrusive spoofing detection.

In recent publications, when a countermeasure is developed, the new approach is presented and evaluated using one single database. However the question whether the countermeasure is really effective in a real world scenario still remains.

This paper addresses this question in two ways. Firstly

[1]http://www.itproportal.com/2011/11/14/ice-cream-sandwich-facial-recognition-cracked/

we present how some countermeasures, reported in the literature, behave in a experimental setup that emulates a more realistic scenario. The designed setup consists in training and tuning countermeasures with one face anti-spoofing database and testing with another one. Secondly we introduce two approaches to train a face anti-spoofing countermeasure to improve its effectiveness in the designed setup.

The remainder of the paper is organized as follows: Section 2 briefly review the relevant literature. Section 3 describes the architecture of the studied countermeasures. Section 4 presents the two publicly available databases which are used in this work. Section 5 discusses the proposed work. The experimental setup and results are discussed in Section 6. Finally, Section 7 summarizes this work highlighting its main contributions.

2. Related work

Considering the type of countermeasures for face anti-spoofing that do not require user collaboration, Chakka et al. in [11] proposed a classification scheme based on the following cues:

- Presence of vitality (liveness);
- Differences in motion patterns;
- Differences in image quality assessment.

Presence of vitality or liveness detection consists of the search of features that only live faces possess. For instance, Pan et al. in [12] exploited the observation that humans blink once every 2-4 seconds and proposed an eye blink-based countermeasure. Experiments carried out with the ZJU Eye Blink Database[2] showed an accuracy of 95.7%.

The countermeasures based on differences in motion patterns rely on the fact that real faces display different motion behavior compared to a spoof attempt. Kollreider et al. [13] present a motion based countermeasure that estimates the correlation between different regions of the face using optical flow field. In this approach, the input is considered a spoof if the optical flow field on the center of the face and on the center of the ears present the same direction. The performance was evaluated using the subset "Head Rotation Shot" of the XM2VTS database whose real access were the videos of this subset and the attacks were generated with hard copies of those data. Using this database, which was not made publicly available, an Equal Error Rate (EER) of 0.5% was achieved. Anjos et al. [14] present a motion based countermeasure measuring the correlation between the face and the background through simple frame differences. Using the PRINT ATTACK database, that approach presented a good discrimination power ($HTER$ equals to 9%).

Countermeasures based on differences in image quality assessment rely on the presence of artifacts intrinsically present at the attack media. Such remarkable properties can be originated from media quality issues or differences in reflectance properties of the object exposed to the camera. Li et al. [15] hypothesize that fraudulent photographs have less high frequency components than real ones. To test this hypothesis a small database was built with 4 identities containing both real access and printed photo attacks. With this private database, an accuracy of 100% was achieved. Assuming that real access images concentrate more information in a specific frequency band, Zhang et al. [9] and Tan et al. [8] used, as countermeasure, a set of Difference of Gaussian filters (DoG) to select a specific frequency band to discriminate attacks and non attacks. Evaluations carried out with the CASIA Face Anti-Spoofing Database (CASIA FASD) and NUAA Photograph Imposter Database[3] showed an Equal Error Rate of 17% and an accuracy of 86% respectively.

Because of differences in reflectance properties, real faces very likely present different texture patterns compared with fake faces. Following that hypothesis, Maatta et al. [16] and Chingovska et al. [10] explored the power of Local Binary Patterns (LBP) as a countermeasure. Maatta et al. combined 3 different LBP configurations ($LBP^{u2}_{8,2}$, $LBP^{u2}_{16,2}$ and $LBP^{u2}_{8,1}$) in a normalized face image and trained a SVM classifier to discriminate real and fake faces. Evaluations carried out with NUAA Photograph Impostor Database [8] showed a good discrimination power (2.9% in EER). Chingovska et al. analyzed the effectiveness of $LBP^{u2}_{8,1}$ and a set of extended LBPs [17] in still images to discriminate real and fake faces. Evaluations carried out with three different databases, the NUAA Photograph Impostor Database, Replay Attack Database and CASIA - FASD [9] showed a good discrimination power with $HTER$ equals to 15.16%, 19.03% and 18.17% respectively.

3. Analyzed Countermeasures

In this paper, we analyze three recently published countermeasures, whose source codes are freely available for download. Figure 1 shows the basic design of the three countermeasures.

After gray-scaling each frame, face detection is applied. The result of the face detection is the input to the countermeasure algorithm which classifies between spoofing attempt and real access. The next subsections presents a brief description of each evaluated countermeasure.

[2] http://www.cs.zju.edu.cn/ gpan/database/db_blink.html

[3] http://parnec.nuaa.edu.cn/xtan/data/NUAAImposterDB.html

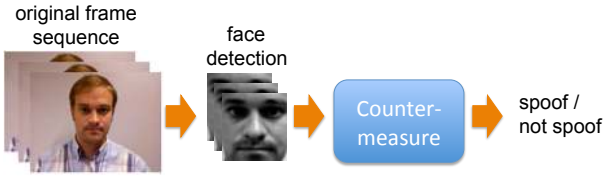

Figure 1. Basic block diagram of the three countermeasures

3.1. Correlation with frame differences

The countermeasure proposed in [14][4] measures the relative motion difference between the face and the background. The authors focused on simple differences of intensities in successive frames. The motion accumulated between this difference (M_D), for a given a Region-of-Interest (RoI) and its respective background, is computed using the following equation:

$$M_D = \frac{1}{S_D} \sum_{(x,y)\in D} |I_t(D) - I_{t-1}(D)| \qquad (1)$$

where D is the RoI, S_D is the area of the RoI and I_t is the intensity of a pixel.

To input the motion coefficient into a classifier, 5 quantities are extracted for every window of 20 frames. The quantities are: the minimum of M_D in that time window, the maximum, the average, the standard deviation and the ratio R between the spectral sum for all non-DC components and DC component itself taken as base the N-point Fourier transform of the signal (see Equation 2). These 5 quantities are fed into a Multi-Layer Perceptron (MLP) classifier with 5 neurons in the hidden layer which is trained to detect spoofing attacks.

$$R = \frac{\sum_{i=1}^{N} |FFT_i|}{|FFT_0|} \qquad (2)$$

3.2. LBP countermeasure

The countermeasure proposed in [10][5] explores the power of Local Binary Patterns (LBP) in static images. The detected faces (see Figure 1) are geometric normalized to 64×64 pixels. The LBP features are extracted from the whole face region and histogrammed. The histograms for each frame are fed into a binary classifier which can be trained to detect spoofing attacks.

Each element of this countermeasure was extensively tuned in [10]. The best configuration reported was $LBP^{u2}_{8,1}$ with Support Vector Machines (SVM) using a Radial Basis Function (RBF) kernel.

[4]http://pypi.python.org/pypi/antispoofing.motion/

[5]http://pypi.python.org/pypi/antispoofing.lbp

3.3. LBP-TOP countermeasure

The countermeasure proposed in [18][6] explored the power of dynamic textures, analyzing motion and texture, in one single descriptor. For that, an approach based on Local Binary Pattern from Three Orthogonal Planes ($LBP-TOP$) was proposed.

The detected faces (see Figure 1) are geometric normalized to 64×64 pixels. After that, $LBP-TOP$ features are extracted from the whole face region and than the histograms for each $LBP-TOP$ plane (XY, XT and YT) are computed. The histograms for each frame are fed into a binary classifier which can be trained to detect spoofing attacks.

Each element of this countermeasure was extensively tuned in [18] and the proposed experimental method resulted in the following configuration: $LBP-TOP^{u2}_{8,8,8,1,1,1}$ with SVM using RBF kernel.

4. Databases

This section presents an overview of the two largest video face spoofing databases. These databases present a set of real access attempts and several fake face attacks of different nature and under varying conditions. To the best of our knowledge, these databases are the only video face spoofing databases publicly available.

The Replay Attack Database (Replay)[7] [10] consists of short video (∼10s) recordings of both real-access and attack attempts to 50 different identities using a laptop. It contains 1200 videos (200 real-access and 1000 attacks) and the attacks were taken in 3 different scenarios with 2 different illumination and support conditions.

The CASIA FASD[8] [9] contains 50 real clients and the corresponding fake faces are captured with high quality from the original ones. The variety is achieved by introducing three imaging qualities (low, normal and high) and three fake face attacks which include warped photo, cut photo (eyeblink) and video attacks. The database contains 600 videos (150 real-access and 450 attacks) and the subjects are divided into subsets for training and testing (240 and 360, respectively).

5. Proposed work

Firstly, we study how the countermeasures, presented in Section 3, will perform in a more realistic condition. This condition consists in training and tuning each one of the countermeasures with one face anti-spoofing database and testing with another one. To report the performance in such a scenario, two evaluation protocols were designed to work

[6]http://pypi.python.org/pypi/antispoofing.lbptop

[7]http://www.idiap.ch/dataset/replayattack

[8]http://www.cbsr.ia.ac.cn/english/FaceAntiSpoof%20Databases.asp

with the databases described in Section 4. These protocols are the "intra-test" protocol and the "inter-test" protocol.

The intra-test protocol is equivalent to the database normal protocol. It consists in training, tuning and testing a countermeasure with the respectively training set, development set and test set of such database. With this protocol, it is possible to evaluate the performance and the generalization power of a countermeasure within one database. The inter-test protocol evaluates the countermeasure performance in a more realistic scenario, close to real usage conditions. It consists in training and tuning a countermeasure with the training set and development set of one database and test it with the test set of another one. With this protocol, it is possible to evaluate the performance and the generalization power of a countermeasure in a set of unseen types of attacks.

The second goal of this work is to provide a way to deal with data from multiple sources (multiple databases) to improve the generalization power of a countermeasure. We introduce two approaches to achieve this goal. The first one is to train a countermeasure with a joint training set combining the train set of multiple databases. The tuning/calibration is done combining the tune set of multiple databases. The performance of this approach is reported using the test set of each database to permit the comparison with intra-test and inter-test experiments.

In the second approach we introduce the Score Level Fusion based Framework. In this framework, each countermeasure is trained independently with the train set of each database. Hereafter, the scores of each countermeasure are merged with a simple sum of normalized scores generating the framework output. The calibration of this framework is done combining the development set of multiple databases. Figure 2 shows the block diagram of the Score Level Fusion based Framework. The performance of this approach is reported using the test set of each database so it is possible to compare the results from the intra-test and inter-test experiments. This technique has the advantage to allow an aggregative approach: it is possible to add more specialized countermeasures to one type of attack.

This work was implemented using the free signal-processing and machine learning toolbox Bob [19] and the source code of the algorithm is available as an add-on package to this framework[9].

5.1. Evaluation protocol

The Replay Attack Database provides a protocol for objective evaluation of a given countermeasure. To mitigate overfitting, such a protocol defines three non-overlapping partitions: the training, development and test set. The training set should be used to train the countermeasure, the development set is used to tune the countermeasure. The test set must be used only to report results.

[9]https://pypi.python.org/pypi/antispoofing.crossdatabase

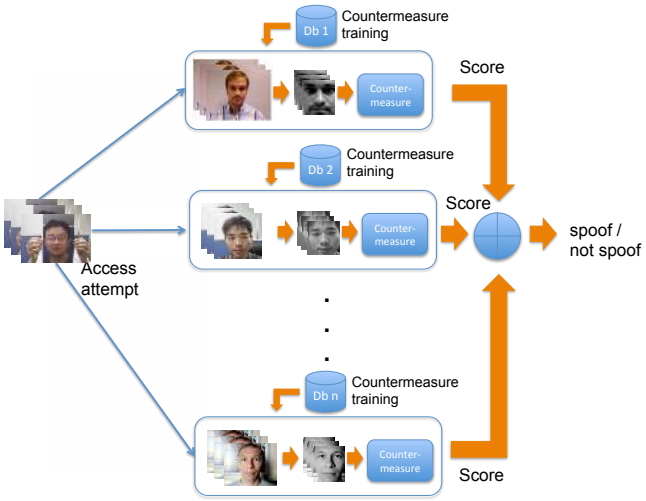

Figure 2. Score Level Fusion based Framework architecture

The CASIA FASD lacks a specific development set; this database has only a train and a test set. To mitigate the overfitting in this database, the train set was split into five partitions and a 5-fold cross-validation training was done. For that, 4 folds were used for training and 1 fold was used to tune the countermeasure. With this strategy, both databases have a train set, a development set (to tune the countermeasure) and a test set, so it is possible to merge them.

The performance of each countermeasure, using the test set of each database, is reported with the Half Total Error Rate ($HTER$):

$$HTER(D_2) = \frac{FAR(\tau(D_1), D_2) + FRR(\tau(D_1), D_2)}{2}, \quad (3)$$

where $\tau(D_n)$ is a threshold, D_n is the dataset, FAR is the False Acceptance Rate in the database D_2 and FRR is the False Rejection Rate in the database D_2. In this protocol, the value of $\tau(D_n)$ is estimated on the Equal Error Rate (EER) using the development set of the database D_1. In this Equation, when $D_1 = D_2$, we have the intra-test protocol and when $D_1 \neq D_2$, we have the inter-test protocol.

Because of 5-fold cross validation procedure, for the CASIA FASD 5 results were generated. The average of $HTER$ was provided as a final result.

In order to eliminate the face detector influence, the same face detector, based on Modified Census Transform (MCT) features [20], was used for both databases.

6. Experiments

This section provides an in-depth analysis of the results obtained with the proposed framework. Three experiments

Table 1. $HTER(\%)$ of each countermeasure applying the intra-test ($D_1 = D_2$) and the inter-test ($D_1 \neq D_2$) protocol.

Countermeasure	**Train/Tune** EER in D_1	**Test** D_2	**HTER(%)** dev	**HTER(%)** test	**HTER degradation (test set) between** $D_1 = D_2$ and $D_1 \neq D_2$
Correlation	Replay $EER = 11.66\%$	Replay	11.66	11.79	309.50%
		CASIA	47.72	48.28	
	CASIA $EER = 26.65\%$	Replay	50.23	50.25	65.68%
		CASIA	26.65	30.33	
$LBPTOP^{u2}_{8,8,8,1,1,1}$	Replay $EER = 8.17\%$	Replay	8.17	8.51	620.68%
		CASIA	60.00	61.33	
	CASIA $EER = 21.59\%$	Replay	48.97	50.64	113.22%
		CASIA	21.59	23.75	
$LBP^{u2}_{8,1}$	Replay $EER = 14.41\%$	Replay	14.41	15.45	274.75%
		CASIA	57.32	57.90	
	CASIA $EER = 24.63\%$	Replay	44.97	47.05	102.89%
		CASIA	24.63	23.19	

were carried out evaluating the performances of the countermeasures using:

1. The intra-test and the inter-test protocol;
2. Combination of multiple databases;
3. Score Level Fusion based Framework.

6.1. Intra-test and inter-test protocol

This experiment analyzes how the three countermeasures perform using the intra-test and the inter-test protocol. Table 1 shows the performance of this experiment following the stablished protocols. The analysis is supported with the ROC curves presented in Figure 3.

Analyzing the performance in the intra-test protocol ($D_1 = D_2$), it can be observed a good performance and a good intra-database generalization power of the three evaluated countermeasures. Note that the countermeasure based on $LBP - TOP$ is the state-of-art in both databases [18] and [21]. The good generalization performance can be attested comparing the results between the development set and the test set. In Table 1 the $HTER(\%)$ in the development set and the $HTER(\%)$ in the test set are very similar. In Figure 3 the ROC curves blue and red (dashed line and solid line) represents the intra-test test protocol. It can be observed that the curves are almost overlapped.

However, analyzing the final performance in the inter-test protocol ($D_1 \neq D_2$), the results considerably degrade compared with the intra-test protocol and it becomes evident that both databases and the methods are strongly biased indicating bad generalization power. The average degradation in all countermeasure is $\sim 247\%$. In Table 1 the $HTER(\%)$ in the development set and the $HTER(\%)$ in the test set are quite different. In Figure 3 the ROC curves blue and green (dotted line and solid line) represents the inter-test test protocol. It can be observed that the curves are quite distant from each other.

The results indicate that the countermeasures and the databases introduce some bias on the spoofing detections. The countermeasures bias are possibly related to the feature selection. The databases bias are possibly related with the types and styles of attacks that is hard to generalize. In next experiment, we stray if the countermeasures are truly biased to databases or can be tuned to overcome the database bias.

6.2. Combination of Multiple Databases

This experiment analyze how the three countermeasures perform using a joint training set combining multiple databases. This is the most intuitive approach to accumulate attacks from different sources. Table 2 shows the performance for each countermeasure trained with this strategy.

Analyzing the performances with this strategy compared with the performance obtained with the inter-set protocol, can be observed a significant improvement for all countermeasures ($\sim 41.5\%$ in HTER average improvement). However, comparing with the intra-test protocol the performance drops drastically ($\sim 62\%$ in HTER average degradation). It can be observed that the performance for CASIA FASD degrades more than for the Replay Attack Database suggesting a strong bias for this database. The Replay Attack Database has twice more data than the CASIA FASD, and this difference is biasing the final performance.

This strategy also has one drawback: when a new database with new types of attacks needs to be added, it is necessary to train and tune all the countermeasures again.

6.3. Score Level Fusion based Framework

To improve the performance results in comparison with the intra-test protocol and the inter-test protocol and to mitigate the bias mentioned in the last section, we introduce a

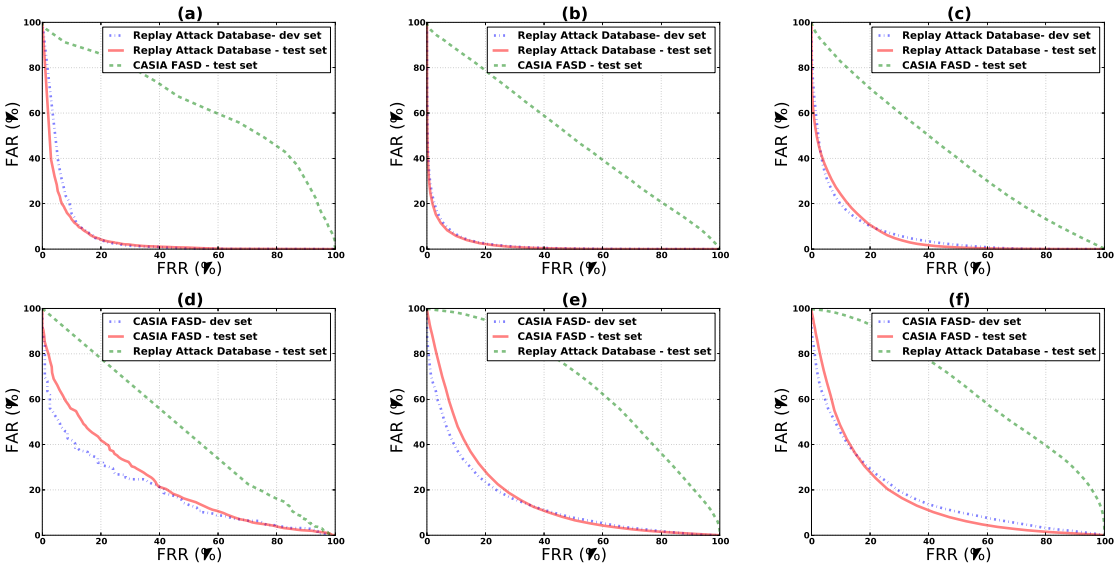

Figure 3. ROC curves of each countermeasure using the intra-test and the inter-test protocol. (a) Correlation with frame differences countermeasure trained and tuned with the Replay Attack Database (b) $LBP-TOP$ countermeasure trained and tuned with the Replay Attack Database (c) LBP countermeasure trained and tuned with the Replay Attack Database (d) Correlation with frame differences countermeasure trained and tuned with the CASIA-FASD (e) $LBP-TOP$ countermeasure trained and tuned with the CASIA-FASD (f) LBP countermeasure trained and tuned with the CASIA-FASD.

Table 2. $HTER(\%)$ of each countermeasure trained with Replay Attack Database and CASIA FASD and test it with each test set of each database.

Countermeasure	**Test**	**HTER(%)** dev	test	**HTER *degradation* (test set) compared with "intra-test" protocol**	**HTER *improvement* (test set) compared with "inter-test" protocol**
Correlation	Replay	12.18	24.14	104.75%	48.03%
	CASIA		43.30	42.76%	10.52%
$LBPTOP^{u2}_{8,8,8,1,1,1}$	Replay	14.29	10.67	25.38%	78.93%
	CASIA		42.04	77.01%	31.45%
$LBP^{u2}_{8,1}$	Replay	20.45	19.07	23.43%	59.47%
	CASIA		45.92	98.01%	20.69%

framework based on score level fusion. Using this framework, when a new countermeasure need to be added, it is possible to "plug it" without any extra training steps required for the other countermeasures.

To support this assumption, we first evaluate the level of independence of the countermeasures trained with different databases in order to ensure its effectiveness in a possible score fusion. Kulcheva and Whitaker [22] show that the combination of statistically independent classifiers is a requirement for a good performance in a score level fusion. In order to evaluate the dependence of classifiers, ten statistics were analyzed. The methodology presented on that work shows that the $Q-statistic$ is most suitable and we choose that metric to evaluate the statistic dependence of each countermeasure for the Score Level Fusion based Framework. The $Q-statistic$ is defined as follow:

$$Q_{R,C} = \frac{N_{11}N_{00} - N_{01}N_{10}}{N_{11}N_{00} + N_{01}N_{10}} \quad (4)$$

where R is the countermeasure trained with the Replay Attack Database; C is the countermeasure trained with CASIA FASD; N_{11} is the number of times that the countermeasure trained with the Replay Attack Database hits (i.e. correctly classifies a sample) and the countermeasure trained with the CASIA FASD also hits; N_{10} is the number of times that the countermeasure trained with the Replay Attack Database hits and the countermeasure trained with the CASIA FASD misses; N_{01} is the number of times that the countermeasure trained with the Replay Attack Database misses and

the countermeasure trained with the CASIA FASD hits and N_{00} is the number of times that the countermeasure trained with the Replay Attack Database misses and the countermeasure trained with the CASIA FASD also misses. The range of this measure goes from -1 to 1.

For statistically independent countermeasures it is expected a $Q_{R,C}$ close to 0. Results close 1 means that both countermeasures are very similar and there is no improvement in the fusion. Results close -1 indicates that both countermeasures oppose each other and a high degradation in the fusion should be expected.

Table 3 shows the statistic dependency using the $Q - statistic$ and the performance in each database trained with the Score Level Fusion based Framework.

Analyzing the $Q - statistic$ it is possible to observe that the Correlation with Frame Differences countermeasure is the most statistically independent and suggests that a score fusion is suitable. This can be attested analysing its performance compared with the inter-test and intra-test protocol results (see Table 1). For the inter-test protocol the improvement with the Score Level Fusion based Framework was significative ($\sim 54\%$ in HTER average improvement). Comparing with the intra-test protocol the degradation was very low ($\sim 5\%$ in HTER average) and the countermeasure is able to detect spoofs in both databases with different degrees of sucess.

However the $Q - statistic$ for the $LBP - TOP$ and the LBP countermeasures present unbalanced values for each database. Specially for the CASIA FASD $Q_{R,C} \simeq -0.4$ suggesting that each one of this two countermeasure trained with different databases oppose each other and are not suitable for the Score Level Fusion based Framework. This can be attested analysing their performances compared with the intra-test protocol results (see Table 1). The degradation is still high ($\sim 102\%$ in $HTER(\%)$ average).

The authors that designed the LBP and $LBP - TOP$ countermeasures chosen the SVM with the RBF kernel as classifier. In both settings, the final trained machines have $\sim 35\%$ of the training data as support vectors, which suggest overfitting in each database. The authors that designed the Correlation with Frame Differences countermeasure chosen MLPs with only 5 neurons, which is much simpler classifier and has less chance to overfit of the training data than a SVM.

It is important to remark that the literature lacks in video face spoofing databases and is not possible to ensure the effectiveness of the Score Level Fusion based Framework in a third database. Its effectiveness in a third video face spoofing database, at this stage is only speculative. Another point to highlight is that the fusion strategy chosen for this work is quite simple. For a future extensions more complex fusion strategies need to be addressed.

7. Conclusion and future work

This article demonstrated how countermeasures developed in the literature perform in a experimental setup that emulates a real world scenario. For that, we introduced two test protocols (inter-test and intra-test protocol) using the only two video face anti-spoofing databases publicly available (Replay Attack Database and CASIA FASD). The evaluation of each countermeasure with the intra-test protocol, suggests a good performance and good intra-database generalization power. However in a more realistic scenario (inter-test protocol) the countermeasures performance degrades $\sim 247\%$ in average of HTER(%). This difference suggests that the countermeasures lacks in generalization and require some improvement.

We studied two approaches to deal with multiple face anti-spoofing databases. The first one combines the train set of each database to train each one of the presented countermeasures. Compared with the inter-test protocol, this strategy improved the countermeasures performance ($\sim 41.5\%$ in HTER average). However, it was observed a strong bias to the Replay Attack Database degrading the performance in the CASIA FASD. In the second approach, we introduced the Score Level Fusion based Framework that merges the scores of countermeasures trained with different databases. Only countermeasures that are statistically independent are suitable for an effective score fusion. Analyzing the $Q - statistic$ measure, the Correlation with Frame Differences countermesure is the most statistically independent and it is the most suitable for the Framework. This was attested comparing the performance of this countermeasure with the performance obtained with the inter-test and intra-test protocols. The HTER average degradation compared with the intra-test protocol was very low ($\sim 5\%$) and the HTER average improvement compared with the inter-test protocol was significative ($\sim 54\%$). However the framework performance using the $LBP - TOP$ and LBP presented unbalanced values for each database and high absolute values for the $Q - statistic$. This behaviour indicated the "improperness" of fusion for these countermeasures. An overfitting caused by the SVM classifier in both countermeasures is a possible reason for this degradation.

As future work we will test more complex strategies of score fusion in order to improve the performance results. This Framework is flexible to aggregate not only data from different databases, but can support any kind of configuration. For example it is possible to aggregate countermeasures trained for a specific kind of attack (print attack, video attack, mobile phone attack and so on). Different configurations for the framework will be tested in the future.

8. Acknowledgments

The research leading to these results has received funding from the European Community's FP7 under grant agree-

Table 3. $Q - statistic$ and $HTER(\%)$ of each countermeasure trained with the Score Level Fusion based Framework and test it with each database.

Countermeasure	**Test**	$Q_{R,C}$	**HTER(%)** dev	test	**HTER degradation (test set) compared with "intra-test" protocol**	**HTER improvement (test set) compared with "inter-test" protocol**
Correlation	Replay	0.11	13.71	12.39	5.09%	75.34%
	CASIA	-0.14		32.08	5.77%	33.55%
$LBPTOP^{u2}_{8,8,8,1,1,1}$	Replay	0.24	23.16	26.04	205.99%	48.58%
	CASIA	-0.41		38.18	60.75%	37.75%
$LBP^{u2}_{8,1}$	Replay	0.38	19.69	21.66	40.19%	53.96%
	CASIA	-0.41		47.16	103.36%	18.55%

ments 257289 (TABULA RASA) and 284989 (BEAT), FUNTTEL / FINEP (Brazilian Telecommunication Technological Development Fund) and CPqD Telecom & IT Solutions.

References

[1] P. Flynn, A. Jain, and A. Ross, *Handbook of biometrics*. Springer, 2008.

[2] S. Li and A. Jain, *Handbook of face recognition*. Springer, 2011.

[3] M. B. Q. Duc, N. M., "Your face is not your password," in *Black Hat conference*, 2009.

[4] U. Uludag and A. Jain, "Attacks on biometric systems: a case study in fingerprints," in *Proc. SPIE-EI*, pp. 622–633, 2004.

[5] J. Leyden, "Gummi bears defeat fingerprint sensors," *The Register*, vol. 16, 2002.

[6] P. Johnson, B. Tan, and S. Schuckers, "Multimodal fusion vulnerability to non-zero effort (spoof) imposters," in *Information Forensics and Security (WIFS), 2010 IEEE International Workshop on*, pp. 1–5, IEEE, 2010.

[7] M. Kanematsu, H. Takano, and K. Nakamura, "Highly reliable liveness detection method for iris recognition," in *SICE, 2007 Annual Conference*, pp. 361–364, IEEE, 2007.

[8] X. Tan, Y. Li, J. Liu, and L. Jiang, "Face liveness detection from a single image with sparse low rank bilinear discriminative model," *Computer Vision–ECCV 2010*, pp. 504–517, 2010.

[9] Z. Zhang, J. Yan, S. Liu, Z. Lei, D. Yi, and S. Li, "A face antispoofing database with diverse attacks," in *Biometrics (ICB), 2012 5th IAPR International Conference on Biometrics*, pp. 26–31, IEEE, 2012.

[10] I. Chingovska, A. Anjos, and S. Marcel, "On the effectiveness of local binary patterns in face anti-spoofing," in *IEEE BIOSIG 2012*, Sept. 2012.

[11] M. Chakka, A. Anjos, S. Marcel, and R. Tronci, "Competition on counter measures to 2-d facial spoofing attacks," in *Proceedings of IAPR IEEE International Joint Conference on Biometrics (IJCB), Washington DC, USA*, Oct. 2011.

[12] G. Pan, L. Sun, Z. Wu, and S. Lao, "Eyeblink-based anti-spoofing in face recognition from a generic webcamera," in *Computer Vision, 2007. ICCV 2007. IEEE 11th International Conference on Computer Vision*, pp. 1–8, IEEE, 2007.

[13] K. Kollreider, H. Fronthaler, and J. Bigun, "Non-intrusive liveness detection by face images," *Image and Vision Computing*, vol. 27, no. 3, pp. 233–244, 2009.

[14] A. Anjos and S. Marcel, "Counter-measures to photo attacks in face recognition: a public database and a baseline," in *International Joint Conference on Biometrics 2011*, Oct. 2011.

[15] J. Li, Y. Wang, T. Tan, and A. Jain, "Live face detection based on the analysis of fourier spectra," *Biometric Technology for Human Identification*, vol. 5404, pp. 296–303, 2004.

[16] J. Maatta and, A. Hadid, and M. Pietikaandinen, "Face spoofing detection from single images using texture and local shape analysis," *Biometrics, IET*, vol. 1, pp. 3 –10, march 2012.

[17] J. Trefnỳ and J. Matas, "Extended set of local binary patterns for rapid object detection," in *Proceedings of the Computer Vision Winter Workshop*, vol. 2010, 2010.

[18] T. de Freitas Pereira, A. Anjos, J. M. De Martino, and S. Marcel, "LBP-TOP based countermeasure against facial spoofing attacks," in *International Workshop on Computer Vision With Local Binary Pattern Variants - ACCV*, Nov. 2012.

[19] A. Anjos, L. E. Shafey, R. Wallace, M. Günther, C. McCool, and S. Marcel, "Bob: a free signal processing and machine learning toolbox for researchers," in *20th ACM Conference on Multimedia Systems (ACMMM), Nara, Japan*, ACM Press, Oct. 2012.

[20] B. Froba and A. Ernst, "Face detection with the modified census transform," in *Automatic Face and Gesture Recognition, 2004. Proceedings. Sixth IEEE International Conference on*, pp. 91–96, IEEE, 2004.

[21] H. A. P. M. Komulainen, J., "Face spoofing detection using dynamic texture," in *International Workshop on Computer Vision With Local Binary Pattern Variants - ACCV*, Nov. 2012.

[22] L. I. Kuncheva and C. J. Whitaker, "Measures of diversity in classifier ensembles and their relationship with the ensemble accuracy," *Mach. Learn.*, vol. 51, pp. 181–207, May 2003.