

Tiago de Freitas Pereira

A COMPARATIVE STUDY OF COUNTERMEASURES TO DETECT SPOOFING ATTACKS IN
FACE AUTHENTICATION SYSTEMS

Campinas
2013

Universidade Estadual de Campinas
Faculdade de Engenharia Elétrica e de Computação

Tiago de Freitas Pereira

A COMPARATIVE STUDY OF COUNTERMEASURES TO DETECT SPOOFING ATTACKS IN
FACE AUTHENTICATION SYSTEMS

Qualificação de Mestrado apresentada na Faculdade de Engenharia Elétrica e de Computação como parte dos requisitos exigidos para a obtenção do título de Mestre em Engenharia Elétrica. Área de concentração: Computação

Orientador: Professor Doutor José Mario De Martino

Este exemplar corresponde a versão final do exame de qualificação apresentado pelo aluno, e orientado pelo Prof. Dr. José Mario De Martino

Campinas
2013

FAÇA AQUI SUA DEDICATÓRIA.

Acknowledgment

Text

A maravilhosa disposição e harmonia do universo só pode ter tido origem segundo o plano de um Ser que tudo sabe e tudo pode. Isto fica sendo a minha última e mais elevada descoberta.

Isaac Newton

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. Countermeasures have been proposed in order to mitigate this vulnerabilities. However several works in the literature present evaluations using different metrics and in private database making the comparison of countermeasures a difficult task. The main goal of this masters project is to provide a comparative study of countermeasures against face *spoofing* attacks.

Key-words: Antispoofing, Liveness detection, Countermeasure, Face Recognition, Biometrics

List of Figures

1.1	2
1.2	Creating a fake fingerprint	3
2.1	Biometric data flow in a face authentication system	8
2.2	New google face unlock screen	9
2.3	Selection of face regions	9
2.4	Block diagram of the countermeasure based on LBP	11
2.5	Block diagram of the DoG countermeasure	12
2.6	Printed photo attacks of the NUAA database	13
2.7	Some frames of real access and spoofing attempts	14
2.8	Example images of real accesses and the corresponding spoofing attempts	15
3.1	(a) Three planes intersecting one pixel (b) LBP histogram of each plane (c) Concatenating the histograms).	18
3.2	Example sequence of a warped photo attack from the CASIA Face Anti-Spoofing Database	20
3.3	Block diagram of the proposed countermeasure.	20
3.4	Face detection strategy for $R_t = 1$	21
3.5	Evaluation of HTER(%) in each plane when the multiresolution area (R_t) is increased with LBP-TOP $_{8,8,8,1,1,R_t}^{u2}$ and LDA classifier - test-set (a) Replay-Attack Database (b) CASIA Face Anti-Spoofing Database.	23
3.6	Evaluation of HTER(%) with LBP-TOP $_{8,8,8,1,1,R_t}^{u2}$ using different classifiers (a) Replay-Attack Database (b) CASIA Face Anti-Spoofing Database.	24
3.7	Evaluation of HTER(%) with LBP-TOP $_{8,8,8,1,1,R_t}^{u2}$ using different LBP configurations in the planes with SVM classifier (a) Replay-Attack Database (b) CASIA Face Anti-Spoofing Database.	25
3.8	Evaluation of the histogram size when (R_t) is increased.	25
3.9	Evaluation of HTER% using LBP-TOP $_{8,8,8,1,1,R_t}^{u2}$ with the single resolution and the multiresolution approach using SVM classifier (a) Replay-Attack Database (b) CASIA Face Anti-Spoofing Database.	26

3.10	Access attempt based evaluation of different time window sizes using mean of features and mean of scores with LBP-TOP _{8,8,8,1,1,1} ^{u2} (a) Replay-Attack Database (HTER %) (b) CASIA Face Anti-Spoofing Database (EER %)	28
3.11	Overall performance of LBP-TOP _{8,8,8,1,1,1} ^{u2} using average of features compared to the DoG baseline method and LBP _{8,1} ^{u2} on the CASIA Face Anti-Spoofing Database.	29
3.12	Performance of LBP-TOP _{8,8,8,1,1,1} ^{u2} using average of features compared to the DoG baseline method and LBP _{8,1} ^{u2} under the different protocols of the CASIA Face Anti-Spoofing Database.	30
4.1	Eye blink countermeasure scheme	35
4.2	Example of the cut photo attack in the CASIA-FASD	38
4.3	Differences in the capture process	39
4.4	Examples of bias in the Replay Attack Database	40
4.5	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) <i>LBP – TOP</i> countermeasure trained and tuned with the Replay Attack Database (c) <i>LBP</i> countermeasure trained and tuned with the Replay Attack Database (d) Correlation with frame differences countermeasure trained and tuned with the CASIA-FASD (e) <i>LBP – TOP</i> countermeasure trained and tuned with the CASIA-FASD (f) <i>LBP</i> countermeasure trained and tuned with the CASIA-FASD.	41
4.6	Joint training scheme for countermeasures	41
4.7	Score Level Fusion based Framework schema	42
4.8	ROC curves of each countermeasure trained with the Score Level Fusion based Framework (a) Correlation with frame differences (b) <i>LBP – TOP</i> countermeasure (c) <i>LBP</i> countermeasure.	44
A.1	LBP operator. (a) and (b) The basic LBP operator, where the neighbourhood of each pixel is thresholded and a binary number is obtained. (c) A circular neighbourhood example (with 8 neighbour points and radius 2). The pixel values are bilinearly interpolated whenever the sampling point is not in the center of a pixel.	50
A.2	All uniform patterns for LBP with 8 neighbours	50

List of Tables

2.1	Performance in <i>HTER</i> (%) terms of the LBP countermeasure in three face spoofing databases.	11
2.2	Number of videos in each subset. Numbers displayed as sums indicate the amount of hand-based and fixed support attack available in each subset.	14
3.1	EER (in %) comparison between the DoG baseline method, $LBP_{8,1}^{u2}$ and $LBP-TOP_{8,8,8,1,1,1}^{u2}$ using average of features on the CASIA Face Anti-Spoofing Database.	28
3.2	EER (in %) development of $LBP-TOP_{8,8,8,1,1,1}^{u2}$ using average of features on the CASIA Face Anti-Spoofing Database.	29
3.3	<i>HTER</i> (%) of the best results achieved on the Replay-Attack Database (following the database protocol) comparing with the provided baseline.	31
3.4	<i>EER</i> (%) of the best results achieved on the CASIA Face Anti-Spoofing Database (following the database protocol) comparing with the provided baseline.	31
4.1	<i>HTER</i> (%) of each countermeasure applying the intra-test ($D_1 = D_2$) protocol.	37
4.2	<i>HTER</i> (%) of each countermeasure applying the inter-test ($D_1 \neq D_2$) protocol.	38
4.3	<i>HTER</i> (%) of the trick countermeasure using only the area of the face bounding box applying the intra-test ($D_1 = D_2$) protocol.	39
4.4	<i>HTER</i> (%) of each countermeasure trained with Replay Attack Database and CASIA FASD and test it with each test set of each database.	42
4.5	<i>Q-statistic</i> and <i>HTER</i> (%) of each countermeasure trained with the Score Level Fusion based Framework and test it with each database.	43

Acronyms

Contents

1	Introduction	1
1.1	Scope and Contributions	5
1.2	Organization of the Thesis	5
2	Spoofing Attacks in Face Authentication	7
2.1	The countermeasures	7
2.1.1	Presence of vitality (liveness detection)	8
2.1.2	Scene	10
2.1.3	Differences in image quality assessment	10
2.2	Face Spoofing Databases	12
2.2.1	NUAA	13
2.2.2	Replay-Attack Database	13
2.2.3	CASIA Face Anti-Spoofing Database	15
2.3	Final Remarks	15
3	Developed Countermeasures	17
3.1	LBP based dynamic texture description	17
3.2	Architecture of the countermeasure	20
3.3	Experiments	21
3.3.1	Effectiveness of each <i>LBP – TOP</i> plane individually and in combination	22
3.3.2	Effectiveness of different classifiers	23
3.3.3	Effectiveness of different LBP operators	24
3.3.4	Effectiveness of the multiresolution approach	26
3.3.5	Access attempt based analysis	27
3.3.6	Discussion	29
3.4	Final Remarks	31
4	Comparative Study of Face Antispoofing Countermeasures	33
4.1	Evaluated countermeasures	33
4.1.1	Motion Correlation	33
4.1.2	Textures with <i>LBP</i>	34

4.1.3	Dynamic Textures with <i>LBP – TOP</i>	34
4.1.4	Eye blinks	34
4.2	Evaluation Protocol	35
4.3	Evaluation Metrics	36
4.4	Evaluated data	36
4.5	Experiments	36
4.5.1	Intra-test protocol	36
4.5.2	Inter-test protocol	38
4.5.3	Combination of Multiple Databases	40
4.5.4	Score Level Fusion based Framework	41
4.6	Final Remarks	44
5	Conclusions	47
5.1	Contributions	48
5.2	Future work	48
A	Texture Description with Local Binary Patterns	49
B	Related Publications	53
References		54

Chapter 1

Introduction

In a modern society, the authentication procedure is an important task to protect data and resources (physical or digital). Consisting of a confirmation of a claimed identity, the authentication procedure is the first and most critical task of security step, restricting access to unauthorized users.

Biometrics is the science of recognizing the identity of a person based on their physical attributes and / or behavior, such as face, fingerprints, hand veins, voice or iris (Li & Jain 2011). The use of biometrics in an authentication procedure has some advantages. Naturally, is not possible to forget or transfer a biometric trait and it hardly disappears (perhaps only in case of a seriously accidents). However, biometrics has some drawbacks. Compared with regular authentication systems, such as passwords or tokens, which are precise, biometric authentication systems have probabilistic behavior. It turns out that biometrics hardly has perfect match; therefore, authentication systems have to deal with error rates. These errors rates can vary depending on a number of factors. As an example, our voice can vary drastically when we get sick or when we are under stress and this impacts a speaker authentication system. Aging, illumination, pose and face expressions are classical issues in face authentication systems.

The use of biometrics in our daily lives has grown in the last decade and we can quote several examples of this. The confirmation of an identity in the brazilian electoral process is based on fingerprints. In Brazil, a bank replaced the use of passwords to palm vein authentication in ATMs. The demand for security pushes the market to biometrics. A recent marked research estimates that the overall market for voice and face biometrics is expected to reach nearly US\$3 billions by the end of 2018¹.

A biometric authentication system can be represented with the simple flow chart in Figure 1.1.

Firstly, the biometric trait is captured using a some kind of sensor. Secondly, the captured biometric trait is processed in order to extract the biometric features. When it is in an enrollment procedure, these features will generate a biometric reference, and it will be stored in a database. In an authentication procedure, these features will be compared with the stored biometric reference. It is possible to observe, in the same Figure, that attacks can be done at any point

¹http://www.biometricupdate.com/201307/voice-biometrics-and-how-far-weve-come/?goback=%2Egde_40210_member_258411747

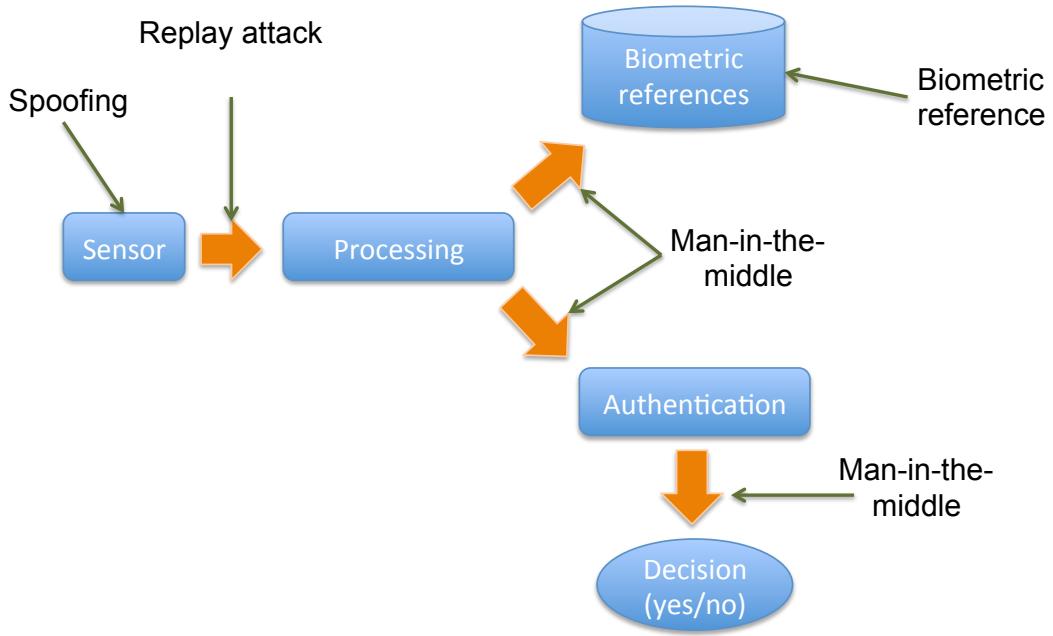


Figure 1.1: Simple architecture of a regular biometric authentication system (adapted from (Xiao 2005))

of the architecture (Xiao 2005).

The **replay attack** is performed by injecting a biometric data previously sent, of the target identity, in order to have a non authorized access. This data can be obtained sniffing the biometric authentication software. To mitigate this kind of attack, the biometric system should ensure that the provided data was not injected artificially (Xiao 2005). A common way to protect against this kind of attack, is to associate a timestamp to the data. As it is improbable to have exactly the same biometric data in different times, this method can be quite effective.

The **biometric reference attack** is performed where the biometrics are stored. This kind of attack, include actions such as the inclusion, removing, modifying and steal biometric references. Among this actions, the possibility to steal a biometric reference is the most dangerous threat, since it is possible to work in a reverse engineering process to regenerate the biometric trait.

Using a hill climbing technique to optimize to the position and the orientation of the minutia (Martinez-Diaz, Fierrez-Aguilar, Alonso-Fernandez, Ortega-Garcia & Siguenza 2006) and (Hill 2001) shown that is possible to generate synthetic fingerprints compatible with fingerprints stored in a database. Fake fingers (with a real fingerprint) with gummy or silicone can be generated with this minutia. It is possible also to inject these minutia in the **Processing** module (see Figure 1.1) in order to deceive the authentication system.

To mitigate the risk of this kind of attack, best practices in security recommends to encrypt the biometric references and to increase the policy to access these biometric references.

In the **man in the middle attack**, the biometric data is intercepted in any point of the architecture in Figure 1.1. As shown in the Figure 1.1, the attacker can:

- Manipulate the matching score;

- Manipulate the biometric authentication response;
- Steal biometric data;
- Inject biometric data (as shown in the replay attack).

The same security recommendations aforementioned to deal with this security breaks can be used here; i.e. encrypt the data before transmission, increase the security grants and so on.

The **spoofing attack**, in biometric systems, is a direct attack to the biometric sensor; i.e. a forged biometric trait is presented to the biometric sensor in order to deceived it. The goal is to pretend to be someone else in order to get forbidden privileges. Most of biometric systems can be spoofed. Next subsections presents a brief discussion about spoofing in different biometric traits:

Fingerprint

In fingerprints verification systems, the attacker can forge a fingerprint with different materials (gummy, silicone, etc). (Matsumoto, Matsumoto, Yamada & Hoshino 2002) and (Leyden 2002) discusses how to generate fake fingerprints using materials easily found in supermarkets. Figure 1.2 shows how easy is to create a molds from a live fingers and to reproduce its fingerprints with gummy. This fake fingers can be used to spoof a fingerprint biometric systems.

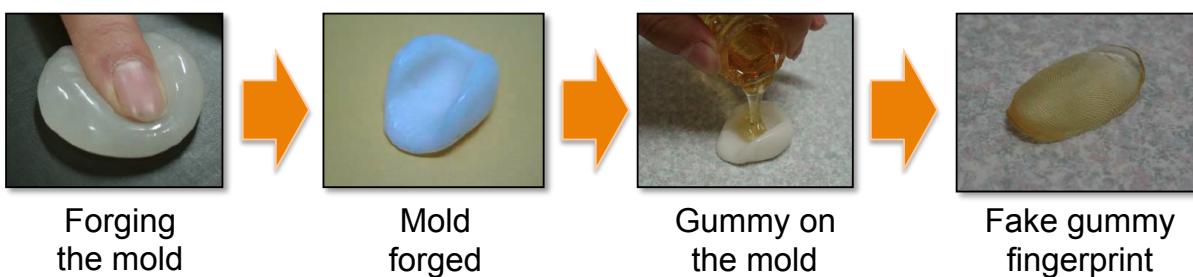


Figure 1.2: Creating a fake fingerprint (Adapted from (Matsumoto et al. 2002))

Recently in Brazil (2013), was reported that doctors in São Paulo were arrested after being caught in the act of using fake fingers made of silicone and imprinted with real fingerprints to defraud a hospital's biometric punch-in clock².

A more sophisticated attack is discussed in (Martinez-Diaz et al. 2006) and (Hill 2001). These papers use a hill climbing technique to optimize the position and the orientation of the minutia. With this optimization was possible to generate fingerprints compatible for match.

Speaker

For speech biometrics, the attacker can forge a human voice by mimicry or recording the voice of the target identity and replaying back to the microphone.

²<http://www.foxnews.com/us/2013/03/13/brazilian-doctors-use-fake-silicone-fingers-to-defraud-hospital/>

(Chetty & Wagner 2004) and (Eveno & Besacier 2005) address the problem using audio-visual features. The first one proposes a bi-modal authentication system using the face information in order to increase security. The second one correlates the lip movements with the content of the speech as a security barrier.

(Zhu, Chatlani & Soraghan 2012) analyse the speech signal itself applying the 1-dimensional *LBP* (Local Binary Pattern) followed by a SVM (Support Vector Machines) in order to detect spoofs.

Iris

Iris biometrics has been traditionally regarded as one of the most reliable and accurate biometric traits, but as the other biometric traits it can also be spoofed. A simple way to spoof an iris recognition system is using a high quality printed image. More sophisticated attacks using contact lenses can also be carried out.

Countermeasures to deal with this kind of attacks can be deployed in hardware (with a specific equipment) or in the software (Galbally, Ortiz-Lopez, Fierrez & Ortega-Garcia 2012). Especially in the software level, (Galbally et al. 2012) addresses the problem using various type of features, including a set of high pass filters, motion features and occlusion filters in the iris images followed by a binary classifier as countermeasure.

An approach based on textures was carried out by (Wei, Qiu, Sun & Tan 2008). This countermeasure uses the co-occurrence matrix descriptor followed by a binary classifier.

Face

Recently, the media has reported 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 (Duc 2009). 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 system³. As a consequence, an eye blinking detection has been introduced in the most recent version of the Android OS. Spoofing in face authentication will be discussed in details in the Chapter 2.

Several technologies related to information security can be deployed in a biometric authentication systems in order to mitigate the mentioned attacks. We can highlight:

- Encrypt the biometric data;
- Improve the security policies;
- Convey the biometric data using a secure channel;

³<http://www.itproportal.com/2011/11/14/ice-cream-sandwich-facial-recognition-cracked/>

- Deploy all modules of the architecture in a physical arrangement that cannot be penetrated;
- Using more than one authentication factor.

However, in a spoofing attack, the target is the biometric sensor, and in the architecture presented in Figure 1.1, is not possible to apply any of the security tools to prevent this kind of attack, becoming the most fragile point. To mitigate this kind of vulnerability, effective countermeasures against face spoofing have to be deployed.

1.1 Scope and Contributions

Focusing in antispoofing countermeasures for face authentication, the goal of this masters dissertation is two fold. The first one, we introduce a novel method to detect face spoofing using the spatiotemporal (dynamic texture) extensions of the Local Binary Pattern. The key idea of the approach is to learn and detect the structure and the dynamics of the facial micro-textures that characterise real faces but not fake ones. The second one, is to provide a comparative study of the state of the art countermeasures for face antispoofing. The key contribution of this comparative study is to covers tests in all video face antispoofing databases freely available focusing in the biases that these databases can introduce in the countermeasures.

1.2 Organization of the Thesis

Besides this introduction, that presented the motivation of this work, the dissertation has four more chapters.

The Chapter 2 defines spoofing attacks in face authentication, presenting the main countermeasures and the resources available for this research.

The Chapter 3 defines and presents the results of the proposed countermeasure based on dynamic texture, the first contribution of this dissertation and their results.

The Chapter 4 defines and presents the results of the comparative study of face antispoofing countermeasures.

Finally the Chapter 5 presents the conclusions and the future work.

Chapter 2

Spoofing Attacks in Face Authentication

As mentioned in the Chapter 1, spoofing attacks in biometrics are direct attacks to the biometric sensor. This chapter discusses the spoofing attacks in face authentication systems and is organized as follows. Section 2.1 discusses spoofing in face biometrics presenting the state of the art countermeasures. Section 2.2 presents the face antispoofting databases publicly available. Finally Section 2.3 presents the final remarks of the chapter.

2.1 The countermeasures

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, aging of subjects and complex outdoor lighting are still research challenges. Advances in the area were extensively reported in (Flynn, Jain & Ross 2008) and (Li & Jain 2011).

It was not until very recently that the problem of spoofing attacks against face biometric system gained attention of the research community. This can be attested by the gradually increasing number of publicly available databases (Pan, Sun, Wu & Lao 2007, Tan, Li, Liu & Jiang 2010, Zhang, Yan, Liu, Lei, Yi & Li 2012, Chingovska, Anjos & Marcel 2012) and contests addressing the problem.

Two contests were organized in the last two years. The first one was organized under IJCB 2011(International Joint Conference on Biometrics)(Chakka, Anjos, Marcel & Tronci 2011) which was the first competition conducted for studying best practices for non-intrusive spoofing detection. More recently, was organized the second competition in this field under ICB 2013 (International Conference on Biometrics)(Chingovska et al. 2013) .

In authentication systems based on face biometrics, spoofing attacks are usually perpetrated using photographs, videos or forged masks. While one can also use make-up or plastic surgery as mean of spoofing, photographs and videos are probably the most common sources of spoofing attacks. Moreover, due to the increasing popularity of social network websites (facebook, flickr, youtube, instagram and others), a great deal of multimedia content - especially videos and photographs - is available on the web that can be used to spoof a face authentication system.

Figure 2.1 (a) and (b) shows the biometric data flow in a real access and in a spoofing attack respectively. In order to mitigate this kind of vulnerability in face authentication systems, effective countermeasures against face spoofing have to be deployed.

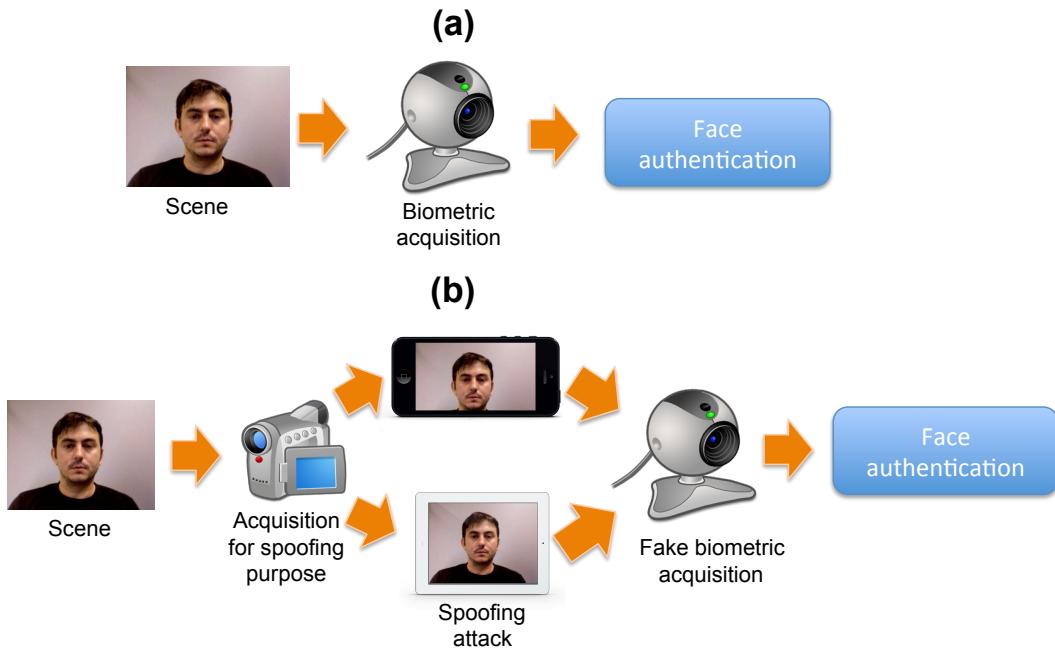


Figure 2.1: (a) Biometric data flow in a real access (b) Biometric data flow in a spoofing attack

The countermeasures against spoofing attempts in face recognition can be macro classified in countermeasures that depend and do not depend on user collaboration. In countermeasures that depends on user collaboration, the user is challenged to interact to the face authentication system. For example, researchers from google are studying a way to unlock the android phones based on facial expressions¹. As can be observed in Figure 2.2, this strategy can be fun in the beginning but in some situations this can be embarrassing. On the other hand, countermeasures that do not depend on user collaboration try to solve this issue analysing the signal itself, without any awareness of the user. This type of countermeasures can be classified by the following cues:

- Presence of vitality (liveness detection);
- Scene characteristics;
- Differences in image quality assessment.

2.1.1 Presence of vitality (liveness detection)

Presence of vitality, or liveness detection, consists of the searching for features that only live faces can possess. The eye blinking is an activity that humans do constantly. A regular human

¹<http://www.bbc.co.uk/news/technology-22790221>



Figure 2.2: New google face unlock screen

blinks once every 2 or 4 seconds in order to maintain the eyes clean and wet. This frequency can vary in stress conditions and/or in a high concentration task. In that situations the interval can extend to ~ 20 seconds. However, does't matter in what condition the person is, the eye blink will always occur. Following that fact, (Pan et al. 2007) propose a countermeasure measuring the eye blinking using Hidden Markov Models (HMM) mapping the state of eyes open and closed. Experiments carried out using a database created by the authors and freely available for download², shown an accuracy of 95.7% .

Supported by the hypothesis that live faces present uncorrelated motion patters in some parts of the face and the attacks do not, (Kollreider, Fronthaler & Bigun 2009) developed a countermeasure based on optical flow field to explore such cue. As a reference to the algorithm, were selected the center of the face an the region of the ears, as can be observed in Figure 2.3.

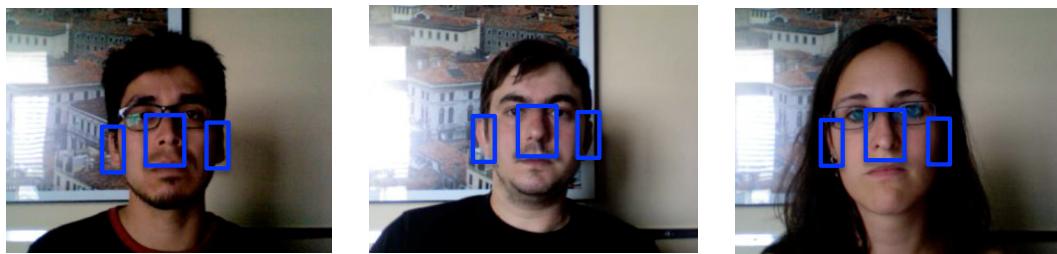


Figure 2.3: Selection of face regions of the algorithm (Kollreider et al. 2009)

The strategy of the countermeasure can be summarized as follows:

1. Detect the face region;
2. Delimitate the region of the face center and the ears (Figure 2.3);
3. Determine if the face region is moving more horizontally or more vertically analysing the optical flow velocities;

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

4. Compute the ratio between the velocities of the delimited areas of the face center and the ears;
5. The spoof is detected if the aforementioned ratio was bigger than a threshold α .

The performance was evaluated using an adaptation of the XM2VTS database. The real accesses were videos from XM2VTS database³ and the attacks were generated with printed photographs from the same database. With this database, which was not made public, an $EER = 0.5\%$ (Equal Error Rate) was achieved.

2.1.2 Scene

Countermeasures that search scene features analyse the relationship of the face in the scene.

The countermeasure proposed in (Anjos & Marcel 2011)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:

$$M_D = \frac{1}{S_D} \sum_{(x,y) \in D} |I_t(D) - I_{t-1}(D)|, \quad (2.1)$$

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 parameters are measured for every window of 20 frames. The parameters 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 2.2). These 10 parameters (five for the face and five for the background) are fed into a Multi-layer Perceptron (MLP) classifier with 5 neurons in the hidden layer which is trained to detect spoofing attacks. This countermeasure was evaluated using the photograph attacks subset of the Replay Attack Database(Chingovska et al. 2012) and achieved an $HTER = 9\%$ (Half Total Error Rate).

$$R = \frac{\sum_{i=1}^N |FFT_i|}{|FFT_0|} \quad (2.2)$$

2.1.3 Differences in image quality assessment

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.

Compared to real faces, attack medias have different reflexive patterns. Supported by that assumption, (Chingovska et al. 2012) and (Maatta and, Hadid & Pietikaandinen 2012), explored

³<http://www.ee.surrey.ac.uk/CVSSP/xm2vtsdb/>

⁴<http://pypi.python.org/pypi/antispoofing.motion/>

the *LBP* (Local Binary Patterns) texture descriptor analysing single frames. In this countermeasure the detected faces (see Figure 2.4) 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.

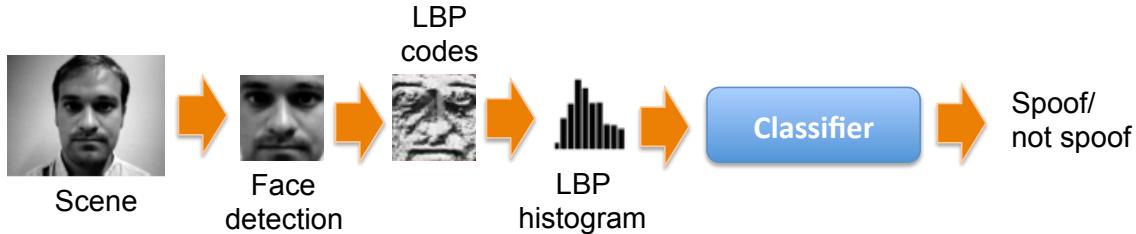


Figure 2.4: Block diagram of the countermeasure based on LBP

The Table 2.1 shows the reported performance, in *HTER* terms, in the three databases; the Replay Attack Database, the CASIA FASD and the NUAA Database using the *SVM* (Support Vector Machines) and *LDA* (Linear Discriminant Analysis) as binary classifiers. In the test set, it can be observed a performance between $\sim 15\%$ and $\sim 20\%$ in the three databases. However, comparing the performance in the development set (used to tune the hyper-parameters) and in the test set of the NUAA database suggest a low generalization capability.

Table 2.1: Performance in *HTER*(%) terms of the LBP countermeasure in three face spoofing databases.

	Replay Attack		NUAA		CASIA-FASD	
	dev set	test set	dev set	test set	dev set	test set
$LBP_{8,1}^{u2} + LDA$	19,60	17,17	0,06	18,32	17,08	21,01
$LBP_{8,1}^{u2} + SVM$	14,84	15,16	0,11	19,03	16,00	18,17

Supported by the assumption that images/videos used in attacks concentrates information in some specifics frequency bands, (Zhang et al. 2012) propose a countermeasure based on Difference of Gaussians filters (DoG).

As can be observed in the block diagram in Figure 2.5, four sequences of DoG filters are applied in the image. Each the gaussian kernel has the size 3×3 and its parameters are:

- $\sigma_1 = 0,5$ e $\sigma_2 = 1$;
- $\sigma_1 = 1$ e $\sigma_2 = 1,5$;
- $\sigma_1 = 1,5$ e $\sigma_2 = 2$;
- $\sigma_1 = 1$ e $\sigma_2 = 2$.

After of the sequence of filters, the images are geometric normalized to 128×128 pixels and these data fed into a *SVM* classifier. Evaluated using the CASIA FASD, the countermeasure

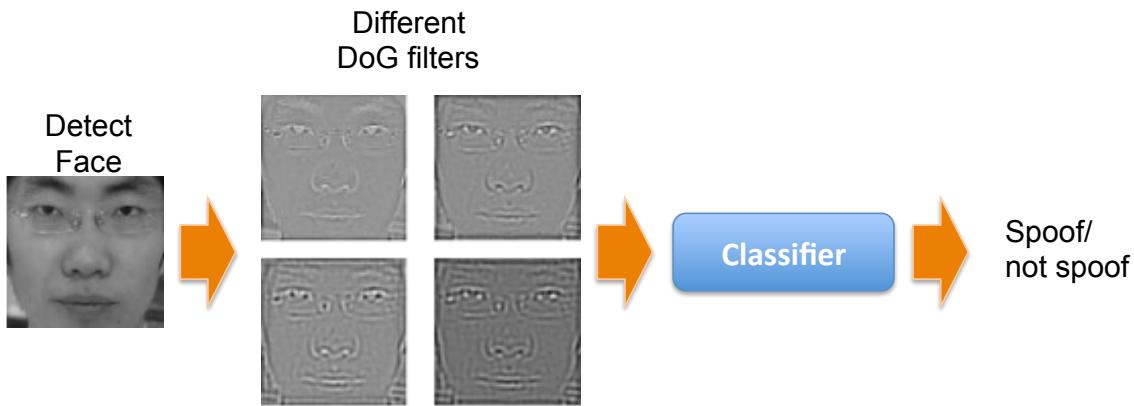


Figure 2.5: Block diagram of the DoG countermeasure

achieved an *EER* of 17%. With these image dimensions, the feature vector of this countermeasure have dimensionality 65,536. The training step of the *SVM* with this dimensionality could take several weeks.

Li et al. (Li, Wang, Tan & Jain 2004) hypothesize that fraudulent photographs have less high frequency components than real ones. To test the hypothesis, a small database was built with 4 identities containing both real access and printed photo attacks. With this private database (which was not made public), an accuracy of 100% was achieved.

In order to detect noise patterns in spoofing attacks, (da Silva Pinto, Pedrini, Schwartz & Rocha n.d.) developed a countermeasure analysing videos combining several elements. First, each frame in a frame sequence is filtered using a gaussian filter followed by a median filter. These filtered images are subtracted by the original ones. The result of this subtraction is so called "residual image". This residual image is analysed in the frequency domain using a 2D Fourier transform. All processed frames in the videos are combined using the Visual Rhythm technique(Zhang, Low, Smoliar & Wu 1995). This technique, generates one image with a combination of all frames ending the preprocessing steps.

A texture description using Gray Level Co-occurrence matrix (GLMC) was applied in the Visual Rhythm image. With the co-occurrence matrix, 12 measures are extracted to fed into a binary classifier that will detect attacks. The classifiers evaluated was the *PLS* (Partial Least Squares) and the *SVM*. With a database combining the photograph subset of the Replay Attack Database and a database created by the author (which was not made public), an AUC (Area Under the Curve) of $\sim 100\%$ was achieved.

2.2 Face Spoofing Databases

In this section, we give an overview of the only three freely available face spoofing databases, the NUAA face antispoofing database, the Replay-Attack Database and the CASIA Face Anti-Spoofing Database (Zhang et al. 2012), consisting of real access attempts and several fake face attacks of different natures under varying conditions.

2.2.1 NUAA

The NUAA face spoofing database⁵ (Tan et al. 2010) consists of images of real accesses and attacks made with printed photographs. Emulating a scenario of access in a regular notebook, this database has images of 15 users split in 3 sections spaced in two weeks. Each section has 4 screenshots per user in different illumination conditions. Figure 2.6 has some examples of this database.



Figure 2.6: Printed photo attacks of the NUAA database

2.2.2 Replay-Attack Database

The Replay-Attack Database⁴ (Chingovska et al. 2012) consists of short video (~ 10 s) 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 three different scenarios with two different illumination and support conditions. The scenarios of attack include:

1. **print**: the attacker displays hard copies of high resolution photographs printed on A4 paper;
2. **mobile**: the attacker displays photos and videos taken with an iPhone 3GS using the phone screen;
3. **highdef**: The attacker displays high resolution photos and videos using an iPad screen with resolution 1024×768 .

The illumination conditions include:

1. **controlled**: the background of the scene is uniform and the light of a fluorescent lamp illuminates the scene;
2. **adverse**: the background of the scene is non uniform and the day-light illuminates the scene.

The support conditions include:

1. **hand-based**: the attacker holds the attack media using his own hands;

⁵<http://parnec.nuua.edu.cn/xtan/data/NuaaImposterdb.html>

Table 2.2: Number of videos in each subset. Numbers displayed as sums indicate the amount of hand-based and fixed support attack available in each subset.

Type	Train	Devel.	Test	Total
Real-access	60	60	80	200
Print-attack	30+30	30+30	40+40	100+100
Mobile-attack	60+60	60+60	80+80	200+200
Highdef-attack	60+60	60+60	80+80	200+200
Total	360	360	480	1200

2. **fixed**: the attacker sets the attack device in a fixed support so it does not move during the spoofing attempt.

Figure. 2.7 show some examples of real accesses and attacks in different scenarios. In the top row, samples from controlled scenario. In the bottom row, samples from adverse scenario. Columns from left to right show examples of real access, printed photograph, mobile phone and tablet attacks.



Figure 2.7: Some frames of real access and spoofing attempts (courtesy of (Chingovska et al. 2012)).

The Replay-Attack database provides a protocol for objectively evaluate a given countermeasure. Such protocol defines three non-overlapping partitions for training, development (tuning) and testing countermeasures. The training set should be used to train the countermeasure, the development set is used to tune the countermeasure and to estimate a threshold value to be used in the test set. The test set must be used only to report results. As performance measurement, the protocol advises the use of Half Total Error Rate (HTER)(Equation 4.2).

$$HTER = \frac{FAR(\tau, D) + FRR(\tau, D)}{2}, \quad (2.3)$$

where τ is the decision threshold, D is the dataset, FAR is the False Acceptance Rate and FRR is the False Rejection Rate. In this protocol, the value of τ is estimated on the Equal Error Rate (EER) using the development set.

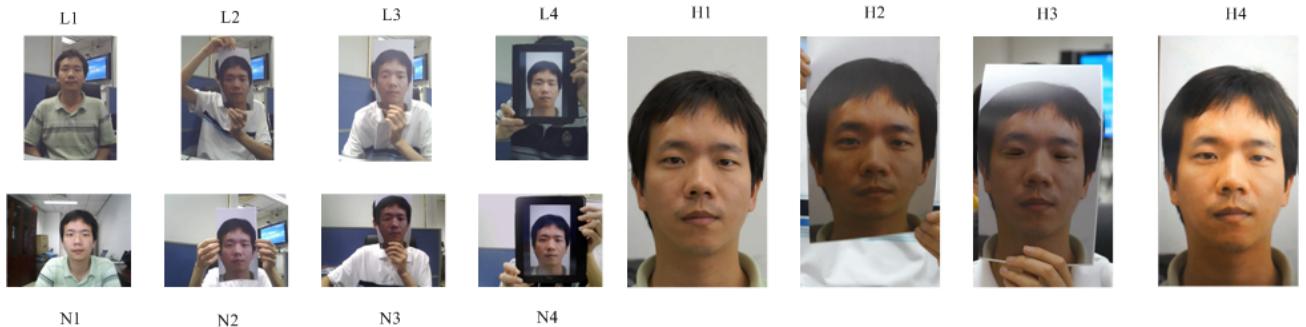


Figure 2.8: Example images of real accesses and the corresponding spoofing attempts (courtesy of (Zhang et al. 2012))

2.2.3 CASIA Face Anti-Spoofing Database

The CASIA Face Anti-Spoofing Database⁵ (Zhang et al. 2012) contains short videos of 50 real clients and the corresponding fake faces were captured with high quality from the original ones. The database has a variety kind of attack. This 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. Examples from the database can be seen in Figure 2.8. Altogether the database consists of 600 video clips and the subjects are divided into subsets for training and testing (240 and 360, respectively). Results of a baseline system are also provided along the database for fair comparison. The baseline system considers the high frequency information in the facial region using multiple DoG features and SVM classifier and is inspired by the work of Tan *et al.* (Tan et al. 2010).

Since the main purpose of the database is to investigate the possible effects of different fake face types and imaging qualities, the test protocol consists of seven scenarios in which particular train and test samples are to be used. The quality test considers the three imaging qualities separately, low (1), normal (2) and high quality (3), and evaluates the overall spoofing detection performance under variety of attacks at the given imaging quality. Similarly, the fake face test assesses how robust the anti-spoofing measure is to specific fake face attacks, warped photo (4), cut photo (5) and video attacks (6), regardless of the imaging quality. In the overall test (7), all data is used to give a more general evaluation. The results of each scenario are reported as Detection-Error Trade-off (DET) curves and equal error rates (EER), which is the point where false acceptance rate (FAR) equals false rejection rate (FRR) on the DET curve.

2.3 Final Remarks

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 (Tan et al. 2010, Zhang et al. 2012, Chingovska et al. 2012) and the recently two contests organized (Chakka et al. 2011, Chingovska et al. 2013). With that efforts, a number of countermeasures were recently published and we presented the

main of them in this chapter.

Most of the countermeasures presented in this chapter were evaluated using different metrics and in some cases in private databases, making the comparison of them a hard task. Just a few of them have source code available, making this task even worse. Additionally, the countermeasures recently published focuses only in the performance analysis in one database. The extension to more databases and the analysis of possible database bias in these countermeasures are overlooked.

Chapter 3

Developed Countermeasures

This chapter presents a countermeasure developed by the author in the scope of this masters dissertation. Micro-texture analysis has been effectively used in detecting photo attacks from single face images (Bai, Ng, Gao & Shi 2010, Maatta and et al. 2012, Chingovska et al. 2012). In this countermeasure the micro-texture analysis is extended to the spatiotemporal domain using the texture descriptor *LBP – TOP* (Local Binary Patterns from Three Orthogonal Planes). The basic theory of Local Binary Patterns in spatiotemporal domain is introduced in Section 3.1. The architecture of the countermeasure is described in Section 3.2. In Section 3.3, we report on the experimental setup and results. Finally, Section 3.4 presents the Final Remarks of the chapter.

The content of this chapter was published in a satellite workshop of the Asian Conference in Computer Vision (ACCV - 2012) with the paper entitled "LBP-TOP based countermeasure against facial spoofing attacks" (Pereira, Anjos, De Martino & Marcel 2012). Additionally this paper was extended and submitted to the journal "EURASIP Journal on Image Processing and Video Processing" organized by Springer with the paper entitled "Face liveness detection using dynamic texture" and is still under revision.

3.1 LBP based dynamic texture description

Määttä et al. (Maatta and et al. 2012) and Chingovska et al. (Chingovska et al. 2012) propose a *LBP* based countermeasures to spoofing attacks based on the hypothesis that real faces present different texture patterns in comparison with fake ones. However, the proposed techniques analyse each frame in isolation, not considering the behaviour over time. As aforementioned, in Chapter 2, motion is a cue explored in some works and in combination with texture can generate a powerful countermeasure. For describing the face liveness for spoofing detection, we considered a spatiotemporal representation which combines facial appearance and dynamics. We adopted the *LBP* based spatiotemporal representation because of its recent convincing performance in modeling moving faces and facial expression recognition and also for dynamic texture recognition (Inen, Pietikäinen, Hadid, Zhao & Ahonen 2011).

The *LBP* texture analysis operator, introduced by Ojala et al. (Ojala, Pietikäinen & Harwood 1996, Ojala, Pietikainen & Maenpaa 2002), is defined as a gray-scale invariant texture

measure, derived from a general definition of texture in a local neighborhood. It is a powerful texture descriptor and among its properties in real-world applications are its discriminative power, computational simplicity and tolerance against monotonic gray-scale changes. The original *LBP* operator forms labels for the image pixels by thresholding a 3×3 neighborhood with the center value and considering the result as a binary number. The histogram of these $2^8 = 256$ different labels is then used as an image descriptor.

The original *LBP* operator was defined to only deal with spatial information. However, more recently it has been extended to a spatiotemporal representation for dynamic texture analysis (DT). This has yielded to the so called Volume Local Binary Pattern operator (*VLBP* (Zhao & Pietikainen 2007)). The idea behind *VLBP* consists of looking at video sequence as a set of volumes in the (X, Y, T) space where X and Y denote the spatial coordinates and T denotes the frame index (time). The neighborhood of each pixel is thus defined in a three dimensional space. Then, similarly to basic *LBP* in spatial domain, volume textons can be defined and organized in histograms. Therefore, *VLBP* combines motion and appearance into a dynamic texture description.

To make *VLBP* computationally treatable and easy to extend, the co-occurrences of the *LBP* on the three orthogonal planes (*LBP – TOP*) was also introduced (Zhao & Pietikainen 2007). *LBP – TOP* consists of the three orthogonal planes: XY , XT and YT , and the concatenation of local binary pattern co-occurrence statistics in these three directions. The circular neighbourhoods are generalized to elliptical sampling to fit to the space-time statistics. The *LBP* codes are extracted from the XY , XT and YT planes, which are denoted as $XY - LBP$, $XT - LBP$ and $YT - LBP$, for all pixels, and statistics of the three different planes are obtained, and concatenated into a single histogram. The procedure is shown in Figure 3.1. In this representation, dynamic texture (DT) is encoded by the $XY - LBP$, $XT - LBP$ and $YT - LBP$.

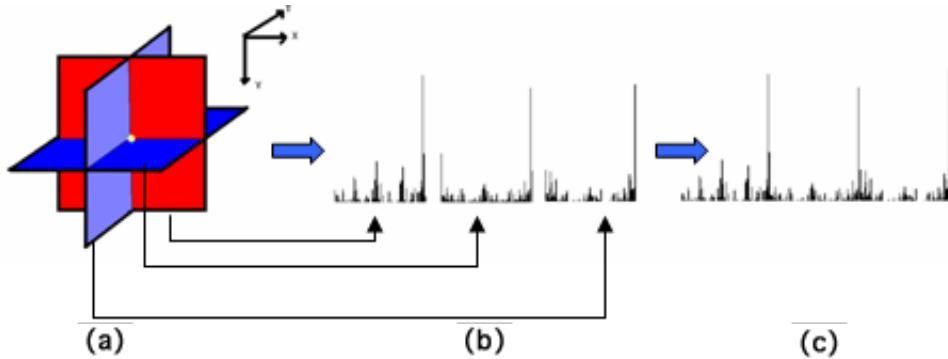


Figure 3.1: (a) Three planes intersecting one pixel (b) LBP histogram of each plane (c) Concatenating the histograms).

Using equal radii for the time and spatial axes is not a good choice for dynamic textures (Zhao & Pietikainen 2007) and therefore, in the XT and YT planes, different radii can be assigned to sample neighbouring points in space and time. More generally, the radii R_x , R_y and R_t respectively in axes X , Y and T , and the number of neighbouring points P_{XY} , P_{XT}

and P_{YT} respectively in the XY , XT and YT planes can also be different. Furthermore, the type of LBP operator on each plane can vary, for example the uniform pattern ($u2$) or rotation invariant uniform pattern ($riu2$) variants (Inen et al. 2011) can be deployed. The corresponding feature is denoted as $LBP - TOP_{P_{XY}, P_{XT}, P_{YT}, R_x, R_y, R_t}^{operator}$.

Assuming we are given a $X \times Y \times T$ dynamic texture

$(x_c \in \{0, \dots, X-1\}, y_c \in \{0, \dots, Y-1\}, t_c \in \{0, \dots, T-1\})$, i.e. a video sequence. An histogram of the DT can be defined as:

$$H_{i,j} = \sum_{x,y,t} I \{ f_j(x, y, t) = i \}, \quad i = 0, \dots, n_j - 1; j = 0, 1, 2 . \quad (3.1)$$

where n_j is the number of different labels produced by the LBP operator in the j_{th} plane ($j = 0 : XY$, $1 : XT$ and $2 : YT$), $f_i(x, y, t)$ expresses the LBP code of central pixel (x, y, t) in the j_{th} plane and I is defined as follows:

$$I(A) = \begin{cases} 1 & \text{if } A \text{ is true} \\ 0 & \text{if } A \text{ is false.} \end{cases} \quad (3.2)$$

Similarly to the original LBP, the histograms are normalized to get a coherent description for comparing the DTs:

$$N_{i,j} = \frac{H_{i,j}}{\sum_{k=0}^{n_j-1} H_{k,j}} . \quad (3.3)$$

In addition to the computational simplification, compared with $VLBP$, $LBP - TOP$ has the advantage to generate independent histograms for each of intersecting planes, in space and time, which can be treated in combination or individually. Because of the aforementioned complexity issues on the implementation of a $VLBP$ based processor, the developed spatiotemporal face liveness description uses $LBP - TOP$ to encode both facial appearance and dynamics.

The key idea of this countermeasure is to learn and detect the structure and the dynamics of the facial micro-textures that characterize real faces but not fake ones. Due to its tolerance against monotonic gray-scale changes, LBP based representation is a large used descriptor for measuring the facial texture quality and determining whether degradations due to recapturing process, e.g. the used spoofing medium, are observed. Instead of just applying static texture analysis, we exploit also several dynamic visual cues that are based on either the motion patterns of a genuine human face or the used display medium.

Unlike photographs and display devices, real faces are indeed non-rigid objects with contractions of facial muscles which result in temporally deformed facial features such as eye lids and lips. Therefore, it can be assumed that the specific facial motion patterns (including eye blinking, mouth movements and facial expression changes) should be detected when a live human being is observed in front of the camera. The movement of the display medium may cause several distinctive motion patterns that do not describe genuine faces. As shown in Figure 3.2 (between the second and the third picture), the use of (planar) spoofing medium might cause sudden characteristic reflections when a photograph is warped or because of a glossy surface of the display medium. As it can be seen, warped photo attacks may cause also distorted facial motion patterns. It is likely that hand-held attacks introduce synchronized shaking of the face and spoofing medium which can be observed as excessive relative motion in the view and facial



Figure 3.2: Example sequence of a warped photo attack from the CASIA Face Anti-Spoofing Database (Zhang et al. 2012) describing the characteristic reflections (flickering) of planar spoofing medium and the distorted motion patterns.

region if the distance between the display medium and the camera is relatively short. Our countermeasure tries to exploit the aforementioned visual cues for face spoofing detection by exploring the dynamic texture content of the facial region. We adopted the *LBP* based spoofing detection in spatiotemporal domain because *LBP – TOP* features have been successfully applied in describing dynamic events, e.g. facial expressions (Zhao & Pietikainen 2007).

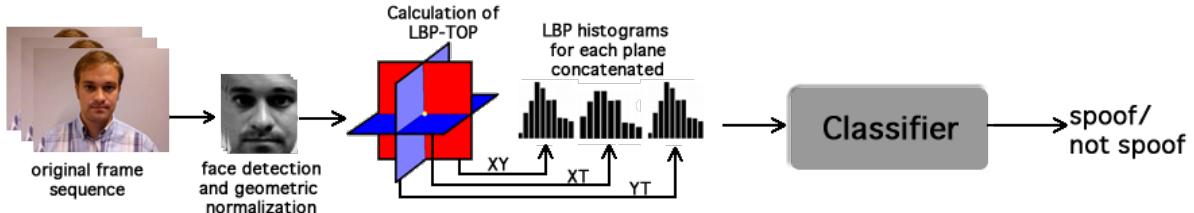


Figure 3.3: Block diagram of the proposed countermeasure.

3.2 Architecture of the countermeasure

Figure 3.3 shows a block diagram of the proposed countermeasure. First, each frame of the original frame sequence was gray-scaled and passed through a face detector using Modified Census Transform (*MCT*) features (Froba & Ernst 2004). Only detected faces with more than 50 pixels of width and height were considered. The detected faces were geometric normalized to 64×64 pixels. The bounding box returned by the automatic face detector introduce some noises in the *LBP – TOP* description. The bounding box, in general, is slightly dislocated in successive frames, even without a translational movement. The *LBP – TOP* descriptor can register movement with this noise. In order to reduce this kind of noise, the same face bounding box was used for each set of frames used in the *LBP – TOP* calculation. As can be seen in the Figure 3.4, the middle frame was chosen. Unfortunately, the face detector is not error free and in case of error in the middle frame face detection, the nearest detection was chosen. If

there is not detected face, in the observed time window, the observation was discarded. After the face detection step, the *LBP* operators were applied for each plane (*XY*, *XT* and *YT*) and the histograms were computed and then concatenated. After the feature extraction step, binary classification can be used to discriminate spoofing attacks from real access attempts.

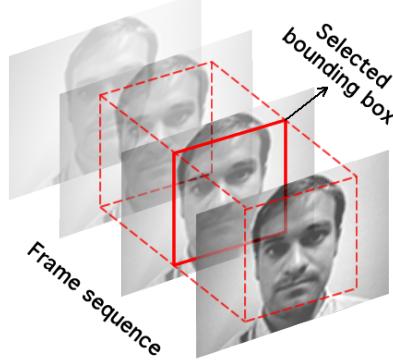


Figure 3.4: Face detection strategy for $R_t = 1$.

Face liveness is rather difficult to be determined based on the motion between couple of successive frames. The used volume can be expanded along the temporal dimension by increasing R_t , as aforementioned in section 3.1. This way to deal with dynamic texture is called single resolution approach, since only one histogram per *LBP – TOP* plane is accumulated. However, this leads to rather sparse sampling on the temporal planes *XT* and *YT*, thus we might loose valuable details. In order to explore the dynamic texture information more carefully, we proposed the multiresolution approach.

The multiresolution approach can be performed by concatenating the histograms in the time domain (*XT* and *YT*) for different values of R_t . The notation chosen to represent these settings is using brackets for the multiresolution data. For example, $R_t = [1 - 3]$ means that the *LBP – TOP* operator will be calculated for $R_t = 1$, $R_t = 2$ and $R_t = 3$ and all resultant histograms will be concatenated. With the multiresolution approach, dense sampling on the temporal planes *XT* and *YT* is achieved.

3.3 Experiments

This section provides an in-depth analysis on the proposed *LBP – TOP* based face liveness description using the Replay-Attack Database (Chingovska et al. 2012) and the CASIA Face Anti-Spoofing Database (Zhang et al. 2012). The *LBP – TOP* representation is computed over relatively short temporal windows and the results are reported using the overall classification accuracy for the individual volumes. Altogether four experiments were carried out evaluating the effectiveness of:

1. Each *LBP – TOP* plane individually and in combination;
2. Different classifiers;

3. Different LBP operators;
4. The multiresolution approach.

In order to study the effect of the different variables, each parameter was tuned solely (fixing other elements) using the development set of each face spoofing database. It should be noted that unlike the Replay-Attack Database, the CASIA Face Anti-Spoofing Database lacks a specific development set. Therefore, the first four experiments were performed in this database using cross-validation by randomly dividing the training data into five folds. Hence, the results presented for CASIA Face Anti-Spoofing Database are actually the average *HTER* on the test set over five iterations of the algorithm with different folds playing the role of a development set.

Finally, we also studied the accumulation of facial appearance and dynamics information over longer time windows and perform an evaluation at system level. The access attempt based results presented in Section 3.3.5 were obtained using the official protocol of each database.

Inspired by (Chingovska et al. 2012), the *LBP-TOP* operator chosen to start the evaluation was $LBP - TOP_{8,8,8,1,1,R_t}^{u2}$.

3.3.1 Effectiveness of each *LBP-TOP* plane individually and in combination

In this experiment, we analysed the effectiveness of each individual plane and their combinations when the multiresolution area is increased. Figure 3.5 shows the *HTER* evolution, on the test set, considering individual and combined histograms of *LBP-TOP* planes for each database. We used, as binary classifier, a linear projection derived from Linear Discriminant Analysis LDA as in (Chingovska et al. 2012).

The results indicate differences in the performance between the two databases. The temporal components (*XT* and *YT*) are a decisive cue for the Replay-Attack Database and the combination of all three planes (*XY*, *XT* and *YT*) gives the best performance. Conversely, for the CASIA Face Anti-Spoofing Database, the addition of temporal planes improves the performance only slightly compared to the spatial *LBP* representation (considering only the *XY* plane). These observations can be explained by taking a closer look at the differences in the databases and their spoofing attack scenarios. 2D fake face attacks can be categorized into two groups, close-up and scenic attacks, based on how the fake face is represented with the spoofing medium.

A close-up spoof describes only the facial area which is presented to the sensor. The main weakness with the tightly cropped fake faces is that the boundaries of the spoofing medium, e.g. a video screen frame, photograph edges, or the attacker's hands are usually visible during the attack, thus can be detected in the scene (Komulainen 2012). However, these visual cues can be hidden by incorporating background scene in the face spoof and placing the resulting scenic fake face very near to the sensor as performed on the Replay-Attack Database. In such cases, the description of facial appearance leads to rather good performance because the proximity between the spoofing medium and the camera causes the recaptured face image to be out-of-focus also revealing other facial texture quality issues, like degradation due to the used spoofing

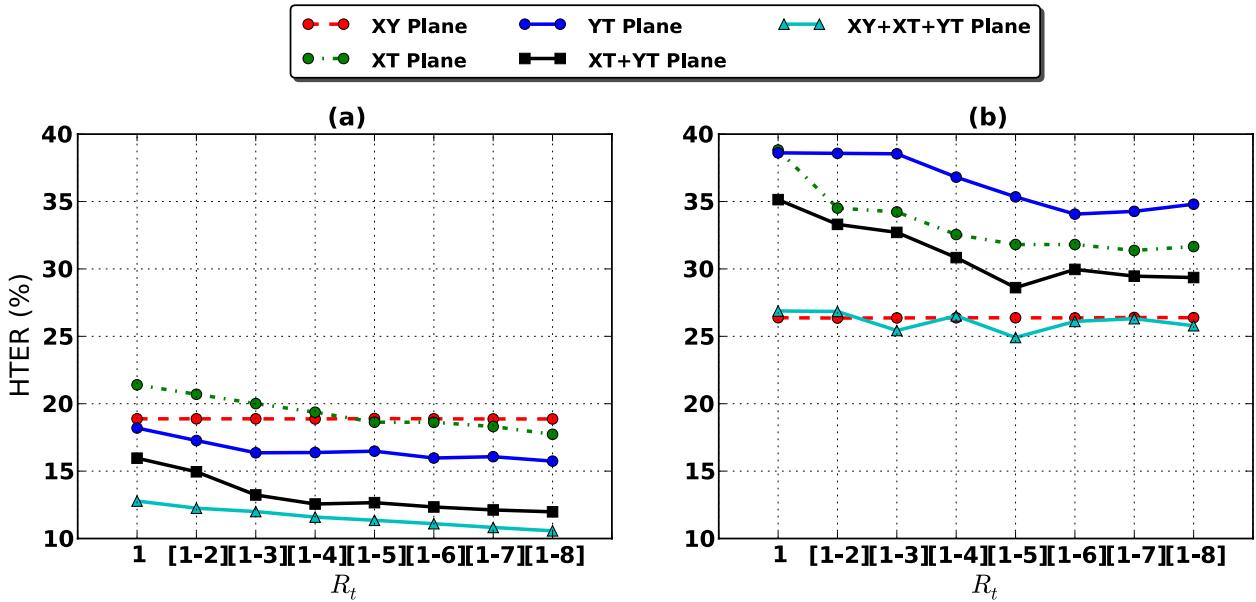


Figure 3.5: Evaluation of HTER(%) in each plane when the multiresolution area (R_t) is increased with LBP-TOP $_{8,8,8,1,1,R_t}^{u2}$ and LDA classifier - test-set **(a)** Replay-Attack Database **(b)** CASIA Face Anti-Spoofing Database.

medium. Furthermore, the attacks in Replay-Attack Database are performed using two types of support conditions, fixed and hand-held. Naturally, the *LBP – TOP* based face representation can easily detect fixed photo and print attacks since there is no variation in the facial texture over time. On the other hand, the hand-held attacks introduce synchronized shaking of the face and spoofing medium. This can be observed as excessive relative motion in the view, again, due to the proximity between the display medium and the sensor. Since the distinctive global motion patterns are clearly visible also on the facial region, they can be captured even by computing the LBP-TOP description over relatively short temporal windows, i.e. low values of R_t .

In contrast, the CASIA Face Anti-Spoofing database consists of close-up face spoofs. The distance between the camera and the display medium is much farther compared to the attacks on Replay-Attack Database. The display medium does not usually move much in the attack scenarios. Therefore, the overall translational movement of a fake face is much closer to the motion of a genuine head. Due to the lack of distinctive shaking of the display medium, the CASIA Face Anti-Spoofing Database can be considered to be more challenging from the dynamic texture point of view. Because the motion cues are harder to explore in some attack scenarios using small values of R_t , we investigated in Section 3.3.5 whether the use of longer time windows helps to reveal the disparities between a genuine face and a fake one.

3.3.2 Effectiveness of different classifiers

In this experiment, we analysed the effectiveness of different classifiers when the multiresolution area is increased. Fig. 3.6 shows the HTER evolution, on the test set, under three different

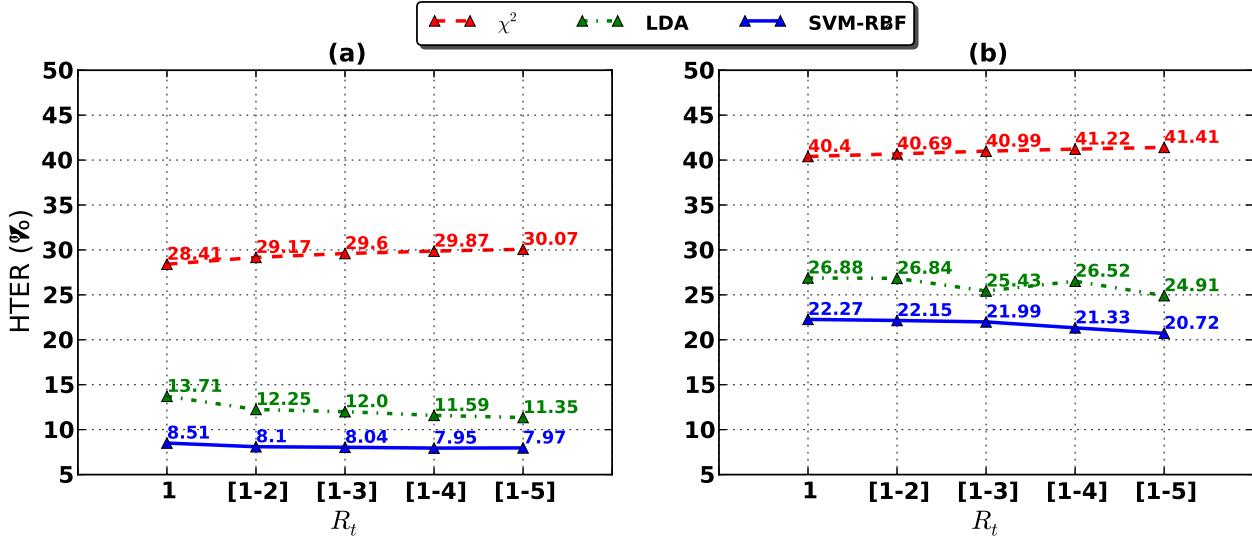


Figure 3.6: Evaluation of HTER(%) with LBP-TOP^{u2}_{8,8,8,1,1,R_t} using different classifiers (a) Replay-Attack Database (b) CASIA Face Anti-Spoofing Database.

classifications schemes. The first one uses χ^2 distance, since the feature vectors are histograms. The same strategy reported in (Chingovska et al. 2012) was carried out. A reference histogram only with real accesses was created averaging the histograms in the training set. The last two selected classification schemes analysed were: Linear Discriminant Analysis (LDA) and Support Vector Machines (SVM) with a radial basis function kernel (RBF).

The SVM classifier with an RBF kernel provided the best performance on the Replay-Attack Database and the CASIA Face Anti-Spoofing Database (7.97% and 20.72% in terms of HTER, respectively). However, it is important to remark that the same LBP-TOP configuration with an LDA classifier resulted in comparable performance (11.35% and 24.91% in terms of HTER). This is not a huge gap and the classification scheme is far simpler. As similar findings have been reported (Chingovska et al. 2012, Komulainen, Anjos, Marcel, Hadid & Pietikäinen 2013), the use of simple and computationally efficient classifiers should be indeed considered when constructing real-world anti-spoofing solutions.

3.3.3 Effectiveness of different LBP operators

The size of the histogram in a multiresolution analysis, in time domain, increases linearly with R_t . The choice of an appropriate LBP representation in the planes is an important issue since it impacts the size of the histograms. Using uniform patterns or rotation invariant extensions, in one or multiple planes, may bring a significant reduction in computational complexity. In this experiment, the effectiveness of different LBP operators in the three LBP-TOP planes (XY , XT and YT) was analysed. Fig. 3.7 shows the performance, in HTER terms, configuring each plane as basic LBP (with 256 bins for $P = 8$), LBP^{u2} (uniform patterns) and LBP^{riu2} (rotation invariant uniform patterns) when the multiresolution area (R_t) is increased in both databases.

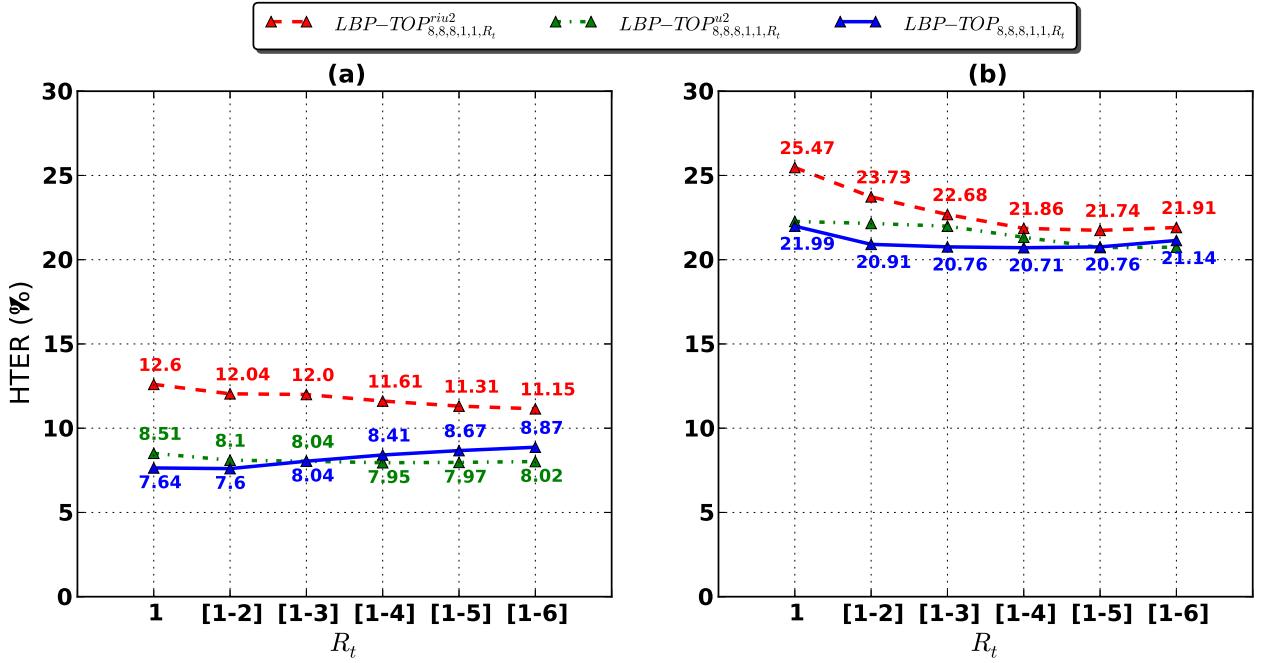


Figure 3.7: Evaluation of HTER(%) with LBP-TOP_{8,8,8,1,1,R_t} using different LBP configurations in the planes with SVM classifier **(a)** Replay-Attack Database **(b)** CASIA Face Anti-Spoofing Database.

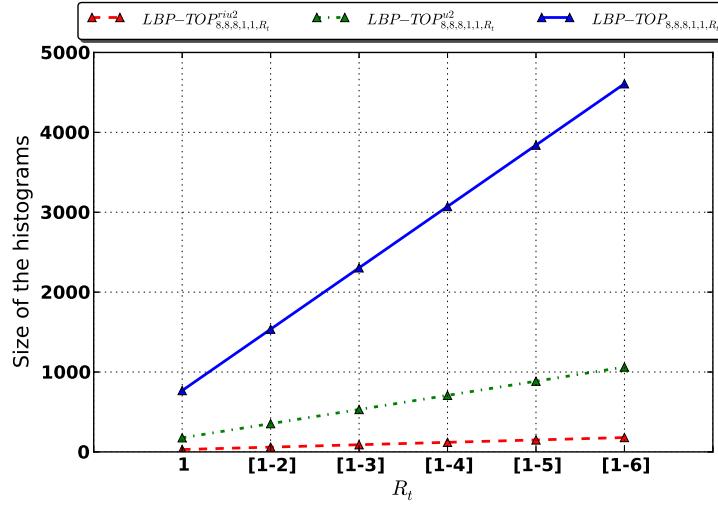


Figure 3.8: Evaluation of the histogram size when (R_t) is increased.

Results must be interpreted with the support of Fig. 3.8, which shows the number of bins on the histograms used for classifications in each configuration.

When the multiresolution area is increased, the HTER saturates for LBP^{riu2} and LBP^{u2} on both datasets. For the basic LBP operator a minimum can be observed in 7.60% and 20.71%

on the Replay-Attack Database and CASIA Face Anti-Spoofing Database respectively. On both databases, basic LBP and LBP^{u2} presented similar performance. Even though the use of regular LBP leads to the best results, the LBP^{u2} operator seems to provide a reasonable trade-off between computational complexity (see Fig. 3.8) and performance. Hence, we will still proceed with LBP^{u2} .

3.3.4 Effectiveness of the multiresolution approach

In this experiment we analysed the effectiveness of the multiresolution approach in comparison to the single resolution approach. The single resolution approach consists of using only fixed values for R_t , without concatenating histograms for each R_t . With this approach the size of the histograms will be constant for different values of R_t , which decreases the computational complexity compared to the multiresolution approach. Fig. 3.9 shows the HTER evolution for different values of R_t in both databases comparing the both approaches.

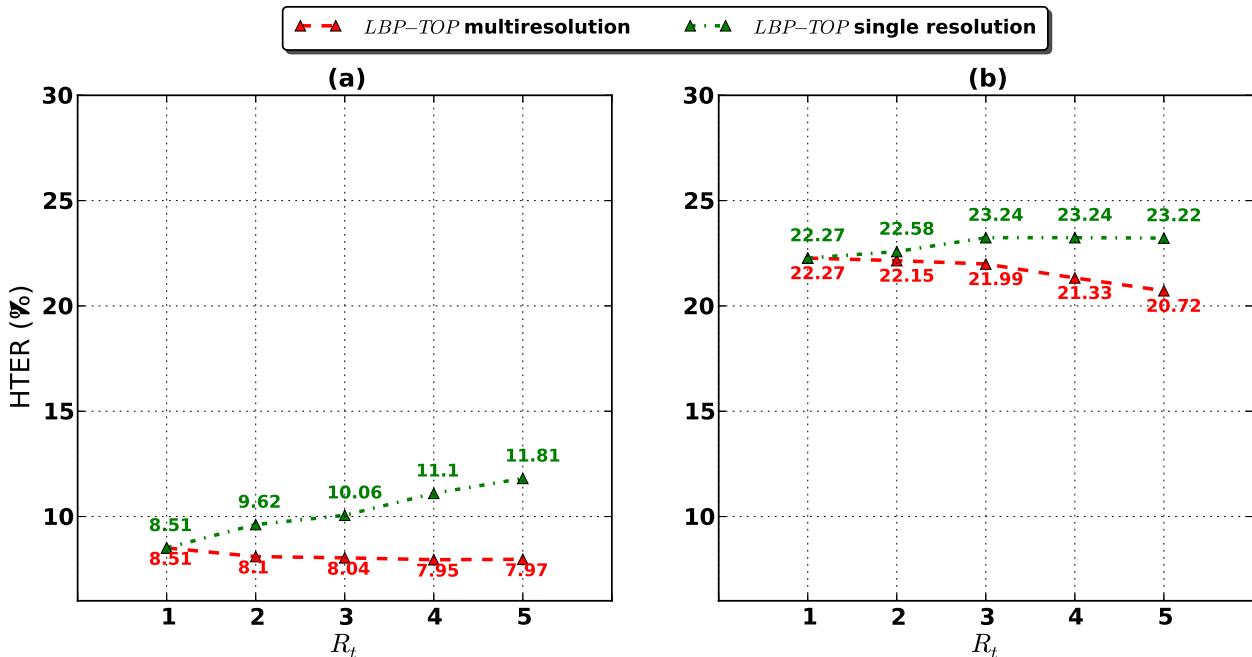


Figure 3.9: Evaluation of HTER% using $LBP-TOP_{8,8,8,1,1,R_t}^{u2}$ with the single resolution and the multiresolution approach using SVM classifier (a) Replay-Attack Database (b) CASIA Face Anti-Spoofing Database.

On both datasets, the HTER of single resolution approach increases with R_t whereas the multiresolution approach helps to keep the HTER low when the multiresolution area is increased. This suggests that the increase of R_t causes more sparse sampling in the single resolution approach when valuable motion information is lost. In contrary, the more dense sampling of the multiresolution approach is able to provide a more detailed description of the motion patterns, thus improving the discriminative power.

3.3.5 Access attempt based analysis

In the previous experiments, the importance of the temporal dimension was studied using the single resolution and the multiresolution approaches. As presented in Section 3.3.1, the multiresolution approach is able to capture the nature of fixed photo attacks and the excessive motion of display medium, especially on the Replay-Attack Database. However, in some attack scenarios, the motion patterns were harder to explore using small values of R_t . We now study how the temporal window size affects the performance when the facial appearance and dynamics information are accumulated over time. The face description of the single resolution and multiresolution methods can be accumulated over longer time periods either by averaging the features within a time window or by classifying each subvolume and then averaging the scores within the current window. In this manner, we are able to provide dense temporal sampling over longer temporal windows without excessively increasing the size of the feature histogram.

In order to follow the method used in previous experiments, we begin evaluating the two averaging strategies with the LBP-TOP $_{8,8,8,1,1,1}^{u2}$ operator and a SVM classifier with RBF kernel. In order to determine the video based system performance, we applied both average of features and scores on the first valid time window of N frames from the beginning of each video sequence. It should be noted that the following access attempt based analysis is based on the official protocol of each database. Thus, the results on Replay-Attack Database are reported in terms of HTER whereas the performance on CASIA Face Anti-Spoofing Database is described using EER.

The access attempt based performance of both averaging strategies on the two databases is presented in Fig. 3.10. The results indicate that when the amount of temporal information increases, the better we are able to discriminate real faces from fake ones. This is the case especially on the CASIA Face Anti-Spoofing Database in which the distinctive motion clues, such as the excessive shaking of the display medium, cannot be exploited. However, when longer video sequences are explored, we are more likely to observe other specific dynamic events, such as different facial motion patterns (including eye blinking, lip movements and facial expression changes) or sudden characteristic reflections of planar spoofing media which can be used for differentiating real faces from fake ones. It is also interesting to notice that by averaging features, more stable and robust spoofing detection performance is achieved on both databases. The averaging scores of individual sub-volumes seems to suffer from outliers, thus more sophisticated temporal processing of scores might lead to more stable behavior.

According to the official test protocol of CASIA Face Anti-Spoofing, also the Detection-Error Trade-off (DET) curves and the EERs for the seven scenarios (Section 2.2.3) should be reported. Based on the previous analysis we chose to use the average of features within a time window of 75 frames which corresponds to three seconds of video time. As it can be seen in Fig 3.11 and Table 3.1, the use of only facial appearance (LBP) leads to better results compared to the baseline method (CASIA baseline). More importantly, when the temporal planes XT and YT are also considered for spatiotemporal face description (LBP-TOP), a significant performance enhancement is obtained (from 16% to 10% in terms of EER), thus confirming the benefits of encoding and exploiting not only the facial appearance but also the facial dynamics information.

More detailed results for each scenario are presented in Fig. 3.12 and in Table 3.1. The

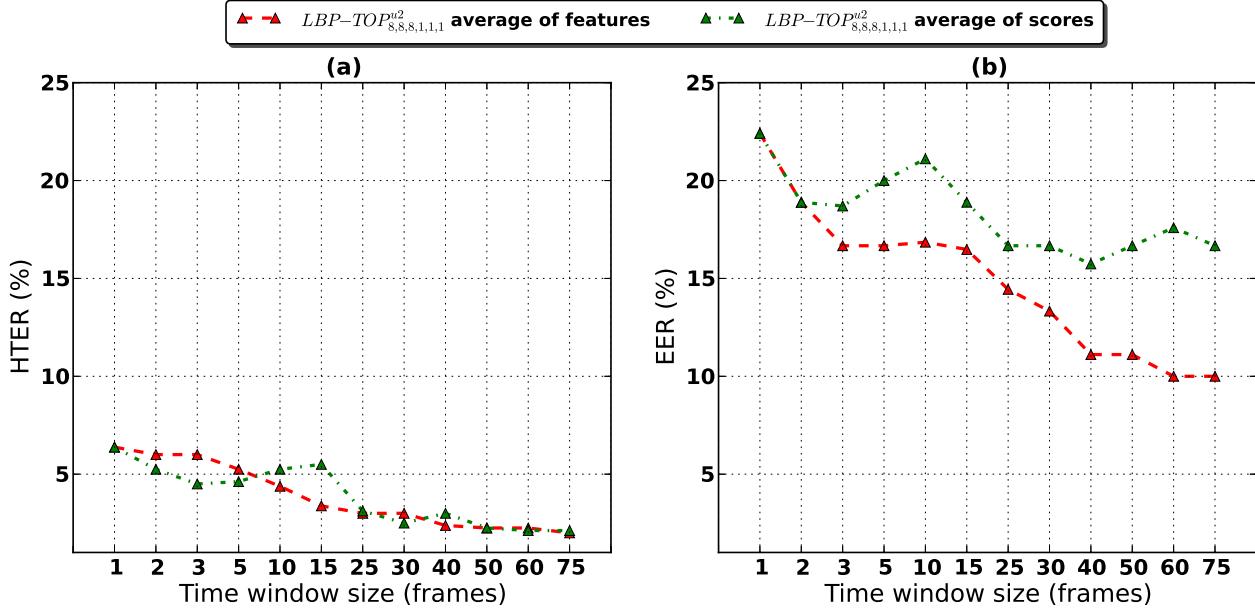


Figure 3.10: Access attempt based evaluation of different time window sizes using mean of features and mean of scores with LBP-TOP^{u2}_{8,8,8,1,1,1} (a) Replay-Attack Database (HTER %) (b) CASIA Face Anti-Spoofing Database (EER %).

Table 3.1: EER (in %) comparison between the DoG baseline method, LBP^{u2}_{8,1} and LBP-TOP^{u2}_{8,8,8,1,1,1} using average of features on the CASIA Face Anti-Spoofing Database.

Scenario	Low	Normal	High	Warped	Cut	Video	Overall
DoG baseline (Zhang et al. 2012)	13	13	26	16	6	24	17
LBP ^{u2} _{8,1}	11	17	13	13	16	16	16
LBP-TOP ^{u2} _{8,8,8,1,1,1}	10	12	13	6	12	10	10

results indicate that the proposed LBP-TOP based face description yields best results in all configurations except under cut-photo attacks. As described in (Zhang et al. 2012), the DoG filtering baseline method is able to capture the less variational nature of the cut eye regions well. However, the difference in the motion patterns seems to be too small for our LBP-TOP based approach as mainly eye blinking occurs during the cut-photo attacks and no other motion is present. The EER development presented in Table 3.2 supports this conclusion since the performance under cut-photo attacks does not improve that much if longer temporal window is applied compared to the other scenarios.

On the other hand, the spatiotemporal face description is able to improve the major drawbacks of DoG based countermeasure. Unlike the baseline method, our approach performs almost equally well at all three imaging qualities. Furthermore, the performance under warped photo and video attacks is significantly better. Especially the characteristic specular reflections (flickering) and excessive and distorted motion of warped photo attacks can be described very well.

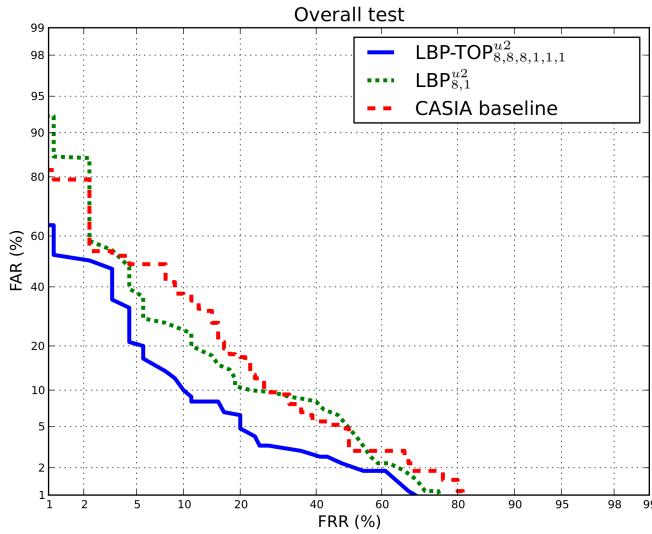


Figure 3.11: Overall performance of LBP-TOP^{u2}_{8,8,8,1,1,1} using average of features compared to the DoG baseline method and LBP^{u2}_{8,1} on the CASIA Face Anti-Spoofing Database.

Table 3.2: EER (in %) development of LBP-TOP^{u2}_{8,8,8,1,1,1} using average of features on the CASIA Face Anti-Spoofing Database.

Frames	Low	Normal	High	Warped	Cut	Video
1	17	27	23	29	16	20
5	13	20	20	19	14	14
10	14	20	19	18	16	14
25	13	13	10	10	14	12
50	13	11	10	7	13	10
75	10	12	13	6	12	10

3.3.6 Discussion

Table 3.3 and Table 3.4 summarize all the results obtained for each database following their provided protocols. In order to be comparable with still frame analysis presented for example in (Chingovska et al. 2012), the results for Replay-Attack Database represent the overall classification accuracy considering each frame individually. The access attempt based results are reported only for CASIA Face Anti-Spoofing Database as requested in its test protocol.

Table 3.3 shows also the results for the LBP⁶ (Chingovska et al. 2012) and the Motion Correlation⁷ (Anjos & Marcel 2011) based countermeasures whose source code is freely available. Table 3.4 contains the provided DoG based baseline and the holistic LBP based face description. It can be seen that the proposed countermeasure presented the best results overtaking the baseline results in both databases, thus confirming the benefits of encoding and exploiting not only the facial appearance but also the facial dynamics information. Unfortu-

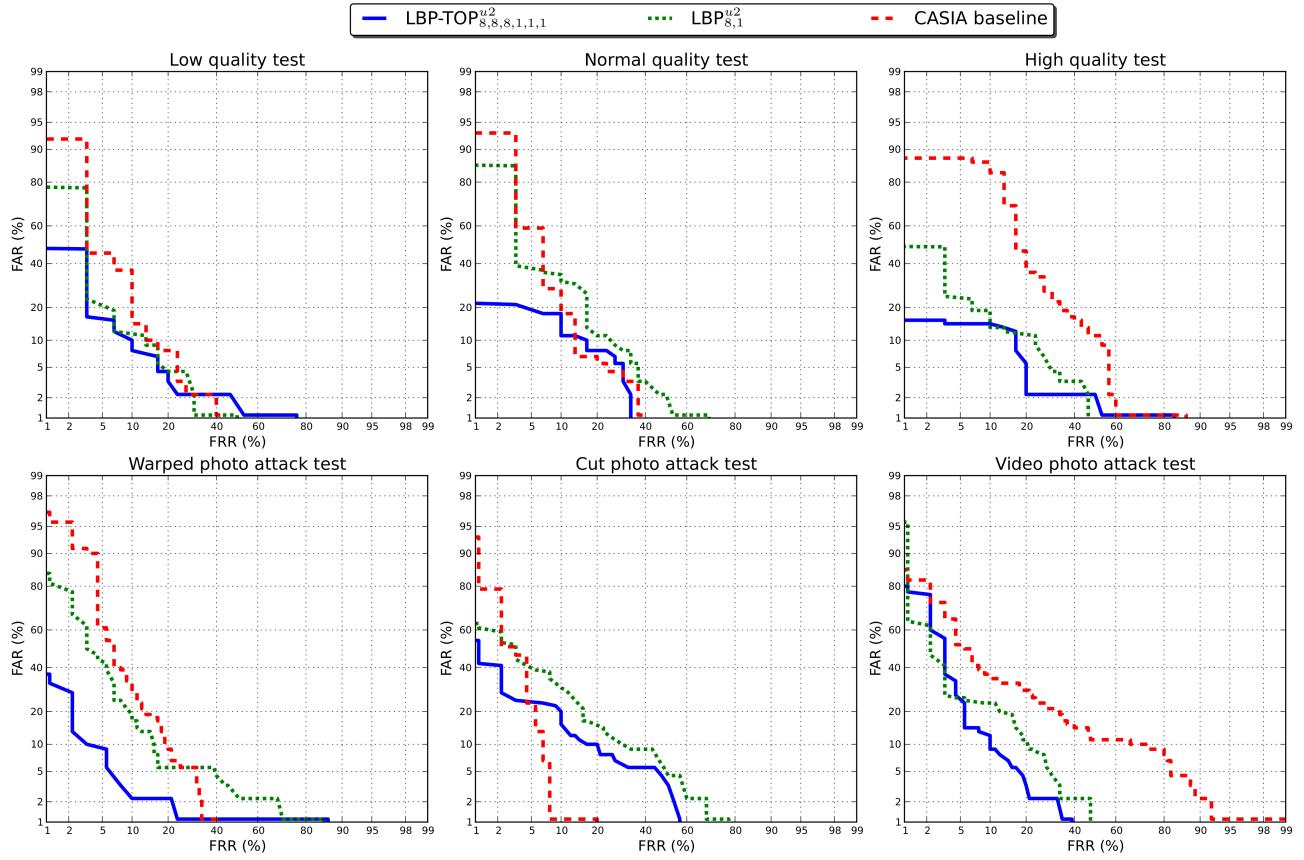


Figure 3.12: Performance of LBP-TOP^{u2}_{8,8,8,1,1,1} using average of features compared to the DoG baseline method and LBP^{u2}_{8,1} under the different protocols of the CASIA Face Anti-Spoofing Database.

nately, our comparison is limited to these countermeasures due to the lack of publicly available implementations of other state-of-the-art techniques presented in literature.

During these experiments we observed that the general performance of the proposed countermeasure was consistently better on Replay-Attack Database compared to the CASIA Face Anti-Spoofing Database. As mentioned in Section 3.3.1, the nature of the attack scenarios is different between the two datasets. In the Replay-Attack Database, our LBP-TOP based face description was able to capture motion patterns of fixed photo attacks and scenic fake face attacks already when only relatively short time windows were explored. Performances below 10% (HTER) were achieved. On the other hand, the CASIA Face Anti-Spoofing Database turned out to be more challenging from the dynamic texture point of view. Due to the lack of motion, analysis of longer temporal windows was required in order to find out distinctive motion patterns between genuine faces and fake ones. As it can be seen in Table 3.4, by extending the micro-texture based spoofing detection into spatiotemporal domain, an improvement from 16% to 10% in terms of EER was obtained. The results also indicate that the proposed dynamic texture based face liveness description was able to improve the state of the art on both datasets.

Table 3.3: HTER(%) of the best results achieved on the Replay-Attack Database (following the database protocol) comparing with the provided baseline.

	dev	test
Motion Correlation (Anjos & Marcel 2011)	11.78	11.79
LBP _{8,1} ^{u2} + SVM	14.84	15.16
LBP _{3×3} + SVM (Chingovska et al. 2012)	13.90	13.87
LBP-TOP _{8,8,8,1,1,1} ^{u2} + SVM	8.17	8.51
LBP-TOP _{8,8,8,1,1,[1–2]} + SVM	7.88	7.60

Table 3.4: EER(%) of the best results achieved on the CASIA Face Anti-Spoofing Database (following the database protocol) comparing with the provided baseline.

	test
DoG baseline (Zhang et al. 2012)	17
LBP _{8,1} ^{u2} + SVM	16
LBP-TOP _{8,8,8,1,1,1} ^{u2} with average of features + SVM	10

3.4 Final Remarks

Inspired by the recent progress in dynamic texture, the problem of face spoofing detection was investigated in this chapter using spatiotemporal local binary patterns. The key idea of the proposed countermeasures consists of analysing the structure and the dynamics of the micro-textures in the facial regions using *LBP – TOP* features that provides an efficient and compact representation for face liveness description. The experiments carried out with this countermeasure consistently outperform prior works on both datasets. Best results were achieved using nonlinear SVM classifier but it is important to notice that experiments with simpler LDA based classification scheme resulted in comparable performance under various spoofing attack scenarios. Thus, the use of simple and computationally efficient classifiers should be indeed considered when constructing real-world anti-spoofing solutions. The results presented in this chapter is reproducible. The source code with instructions on how to reproduce the results is freely available¹.

¹<https://pypi.python.org/pypi/antispoofing.lbptop/>

Chapter 4

Comparative Study of Face Antispoofing Countermeasures

This chapter provides the experiments and the results of this comparative study of countermeasures. The section 4.1 presents the countermeasures compared in this dissertation and how each hyper-parameter was set. The Section 4.2 presents the evaluation protocol applied in this dissertation and section 4.3 presents metrics used. Section 4.4 presents the data evaluated in this dissertation and how this data was organized in the experiments. In Section 4.5 we report our experimental results. Finally, Section 4.6 presents the final remarks of the chapter.

The content of this chapter was published in the International Conference on Biometrics (ICB 2013) with the paper entitled "Can Face Antispoofing Countermeasures Work in a Real World Scenario?" (de Freitas Pereira, Anjos, De Martino & Marcel 2013).

4.1 Evaluated countermeasures

For this comparative study were selected four countermeasures that do not depend of the user collaboration. Very representative according to the state of the art in the face antispoofing research, each one explore one of the main cues mentioned in the Section ?? (Presence of vitality, Scene characteristics, and Differences in image quality assessment). Next subsections presents the details of each countermeasure and how was set the hyper-parameters of each one.

4.1.1 Motion Correlation

As presented in Section 2.1.2, the Motion correlation (Anjos & Marcel 2011) countermeasure measures the correlation between the face and it background. The source code of this countermeasure is freely available¹ in order to reproduce the results of the published paper. There are, basically, two hyper-parameters in this countermeasure. The first one, is the number of frames used to compute the 5 quantities. The second one is the binary classifier.

As the authors suggested, twenty frames to compute the 5 quantities are sufficient to the algorithm converge in their experiments. The classifier suggested in the paper was one based

¹<https://pypi.python.org/pypi/antispoofing.motion/>

on Multi-layer Perceptron. This classifier has, basically, the number of hidden layers and the number neurons in each hidden layer as hyper-parameters. The authors suggested one hidden layer and five neurons in this hidden layer as good tradeoff between computational complexity and performance. We will adopt the suggested hyper-parameters in our experiments.

4.1.2 Textures with *LBP*

Presented in Section 2.1.3, the countermeasure based on Textures with *LBP* (Chingovska et al. 2012) and (Maatta and et al. 2012) explore the differences in texture properties between real accesses and attacks in single frames. The source code of this countermeasure is freely available² in order to reproduce the results of the published paper.

There are, basically, three hyper-parameters in this countermeasure. The first one is the geometrically normalized face size. The authors suggested a face size of 64×64 pixels. The second one is the configuration of the *LBP* texture descriptor. The *LBP* itself has several hyper-parameters (Inen et al. 2011) and the authors of both papers stressed only some of that. In this dissertation we will follow the setup suggested by (Chingovska et al. 2012) using the $LBP_{8,1}^{u2}$. Finally the last hyper-parameter is the binary classifier. The best classifier tested by (Chingovska et al. 2012) was the Support Vector Machines (SVM) using the Radial Basis Function (RBF).

4.1.3 Dynamic Textures with *LBP – TOP*

Presented in the Chapter 3 as one our contributions, the countermeasure based on dynamic textures with *LBP – TOP*, explore the texture dynamics to detect attacks in a frame sequence. The source code of this countermeasure is freely available³ in order to reproduce the results of the published paper.

There are, basically, three hyper-parameters in this countermeasure. The first one is the geometrically normalized face size. In our previous experiments we worked with face sizes of 64×64 pixels and we will keep this in the next experiments. The second hyper-parameter is the configuration of the *LBP – TOP* descriptor. As in the *LBP*, the *LBP – TOP* descriptor itself has several hyper-parameters and most of it was extensively tuned in Chapter 3. As good tradeoff between computational complexity and performance, we selected the following configuration: $LBP – TOP_{8,8,8,1,1,1}^{u2}$ with a single resolution. The last hyper-parameter is the binary classifier. The evaluation method proposed in the Chapter 3 suggests the SVM classifier with RBF kernel.

4.1.4 Eye blinks

The eye blink countermeasure used in this dissertation, uses a similar technique applied in Motion Correlation countermeasure(Anjos & Marcel 2011). The difference is; the accumulated motion M_D , (see Equation 2.1) is computed between the face region and the eyes region as can be observed in the Figure 4.1.

²<https://pypi.python.org/pypi/antispoofing.lbp/>

³<https://pypi.python.org/pypi/antispoofing.lbptop/>



Figure 4.1: Eye blink countermeasure scheme. The eye blink is measured as a motion correlation between the eyes region and the face region.

The eye blink score for each single frame n in a frame sequence, is computed using the following equation:

$$S_n = \frac{M_{D_{eye}}(n)}{M_{D_{face}}(n)} - ravg\left(\frac{M_{D_{eye}}}{M_{D_{face}}}\right)(n) \quad (4.1)$$

where the $ravg$ is the remainder average in a frame sequence until the frame n .

The trigger of this countermeasure is the number of blinks. In this dissertation, we will test one, two and three blinks as a trigger of a real access. The source code of this countermeasure is freely available⁴.

4.2 Evaluation Protocol

For this comparative study, we will evaluate the intra-database and the inter-database (or cross-database) generalization. For that, we developed two test protocols, the intra-test protocol and the inter-test protocol.

The intra-test protocol evaluates the intra-database generalization. It consists in training, tuning and testing a countermeasure with the respectively training set, development set and test set of one database.

The inter-test protocol is a little bit more challenging, since test the inter-database generalization (or cross-database). 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 others databases.

⁴<https://github.com/bioidiap/antispoofing.eyeblink>

4.3 Evaluation Metrics

The final performance of each countermeasure using both evaluation protocols in the test set of each database is reported with the Half Total Error Rate (*HTER*):

$$HTER(D_2) = \frac{FAR(\tau(D_1), D_2) + FRR(\tau(D_1), D_2)}{2}, \quad (4.2)$$

where $\tau(D_1)$ is the decision 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, to measure the performance using the intra-database protocol, is necessary to consider $D_1 = D_2$. To measure the performance using the inter-database protocol, just consider $D_1 \neq D_2$.

4.4 Evaluated data

As the Motion correlation, *LBP – TOP* and the eye blink countermeasures need a frame sequence to work, the databases evaluated in this dissertation will be the Replay Attack Database (Section 2.2.2) and CASIA Face Antispoofing Database (Section 2.2.3).

As already mentioned in Section 2.2.2 the Replay Attack Database has three non-overlapping partitions; the training, development and test set for respectively train, tune and test a countermeasure. To run the proposed protocols in this database, we will use the train set to train the four countermeasures; the development set will be used to estimate the value of $\tau(D_1)$. Finally the test set will be used to report the *HTER*(D_2).

The CASIA FASD lacks a specific development set; this database has only a train and a test set. Since we need the three sets (train, development and test), we split the train set in 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 estimate the value of $\tau(D_1)$. The original test set was preserved, to report the *HTER*(D_2). Because of 5-fold cross validation protocol for the CASIA FASD, five results were generated. The average of *HTER* was provided as a final result.

4.5 Experiments

4.5.1 Intra-test protocol

Table 4.1 shows the performance of the four countermeasures, in *HTER* terms, applying the Intra-test protocol.

Analyzing the performance in the intra-test protocol ($D_1 = D_2$) it can be observed that different countermeasures have different performances using different databases. As already discussed in the Section 3.3.1, both databases has some differences that impacts in the final performance in each database, making the CASIA-FASD a more difficult database than the

Table 4.1: $HTER(\%)$ of each countermeasure applying the intra-test ($D_1 = D_2$) protocol.

Countermeasure	Train/Tune D_1	Test D_2	$HTER(\%)$		$FAR(\%)$		$FRR(\%)$	
			dev	test	dev	test	dev	test
Correlation	Replay	Replay	11.66	11.79	11.66	10.53	11.66	13.05
	CASIA	CASIA	24.91	31.36	24.91	32.52	24.91	30.21
$LBPTOP^{u2}_{8,8,8,1,1,1}$	Replay	Replay	8.17	8.51	8.17	7.42	8.17	9.60
	CASIA	CASIA	21.77	22.27	21.77	24.24	21.77	20.33
$LBP^{u2}_{8,1}$	Replay	Replay	14.41	15.45	14.41	17.32	14.41	13.63
	CASIA	CASIA	23.00	22.54	23.00	24.78	23.00	20.3
Eye blink (1 blink)	Replay	Replay	48.17	52.62	89.67	90.25	6.67	15.00
	CASIA	CASIA	48.61	48.33	97.22	93.33	0.00	3.33
Eye blink (2 blink)	Replay	Replay	53.50	54.87	10.33	16.00	96.67	93.75
	CASIA	CASIA	41.67	44.81	8.33	6.30	75.00	83.33
Eye blink (3 blink)	Replay	Replay	49.17	49.50	0.00	0.25	98.33	98.75
	CASIA	CASIA	47.22	48.89	2.78	0.00	91.67	97.78

Replay Attack Database. These differences impacted in our proposed countermeasure, based on $LBP - TOP$ (Chapter 3), and it seems to impact in other countermeasures.

The exception here is the countermeasure based on eye blinks. In both databases de performances, in $HTER$ terms, vary from $\sim 40\%$ to $\sim 50\%$ independently of the number of blinks that we consider, which is worse compared to the other three countermeasures.

A closer observation to the FAR and FRR in this countermeasure, we can conclude some things. Considering one blink as liveness check was observed a FAR of $\sim 90\%$ in both databases. Both databases has video attacks and the countermeasure capture eye blinks from there. Specially in the Replay Attack Database, the hand-held attacks introduce some noise that hits the liveness check. CASIA FASD has the warped photo attacks and these warps made by the attacker also introduce some noise deceiving the eye blink system. Also the CASIA FASD has the cut photo attacks, where the attacker uses masks of the target identity with holes in the eyes region, as can be observed in the Figure 4.2. This attacks have a real eye blinks.

Increasing the number of eye blinks (two and three) as a liveness check, in order to increase the robustness, the final performance is still not satisfactory. In $HTER$ terms $\sim 50\%$ in the test set in both databases. The FAR now is close to 0% but the FRR is greater than 93% in both databases. The videos in these databases are short ($\sim 10s$) and it turns out that the people don't blink twice in this short recording window. With these evidences of bad performances, we are no longer to support the eye blink countermeasure in this dissertation.

However, it is possible to observe that the $LBP - TOP$, LBP and Motion correlation countermeasures have a good overall performance and, the most import, a good generalization capability. In Table 4.1 the $HTER$ in the development an in the test set are very similar indicating the assumption of generalization. The ROC curves in Figure 4.5 corroborates this assumption. In the figure, the curves blue and red (dotted line and solid line) represents the



Figure 4.2: Example of cut the photo attack in the CASIA-FASD. It is possible to see the eye blink in third frame.

intra-test test protocol. It can be observed that the curves are almost overlapped.

4.5.2 Inter-test protocol

Table 4.2 shows the performance of the three countermeasures, in *HTER* terms, applying the Inter-test protocol.

Table 4.2: *HTER*(%) of each countermeasure applying the inter-test ($D_1 \neq D_2$) protocol.

Countermeasure	Train/Tune	Test	<i>HTER</i> (%)	
	D_1	D_2	dev	test
Correlation	Replay	CASIA	11.66	61.78
	CASIA	Replay	24.91	48.47
$LBPTOP_{8,8,8,1,1,1}^{u2}$	Replay	CASIA	8.17	51.05
	CASIA	Replay	21.77	61.11
$LBP_{8,1}^{u2}$	Replay	CASIA	46.87	48.06
	CASIA	Replay	23.00	57.64

Analyzing the performance in the inter-test protocol ($D_1 \neq D_2$), it can be observed that the performance results considerably degrade compared with the intra-test protocol and it becomes evident that both databases and the methods are strongly biased indicating that the countermeasures do not generalize as expected. In Table 4.2 the *HTER* in the development set and in the test set are quite different. In Figure 4.5 the ROC curves blue and green (dashed line and solid line) representing the curves got by the development set of the database D_1 and by the test set of the database D_2 when $D_1 \neq D_2$ respectively, are quite distant from each other.

The results indicate that the differences in the databases can bias the countermeasures. Was observed two kinds of database bias. The first one is relative to process of capture of the databases and we call this of **capture bias**. The second one is relative to the differences of attacks in both databases and we call this of **attack bias**.

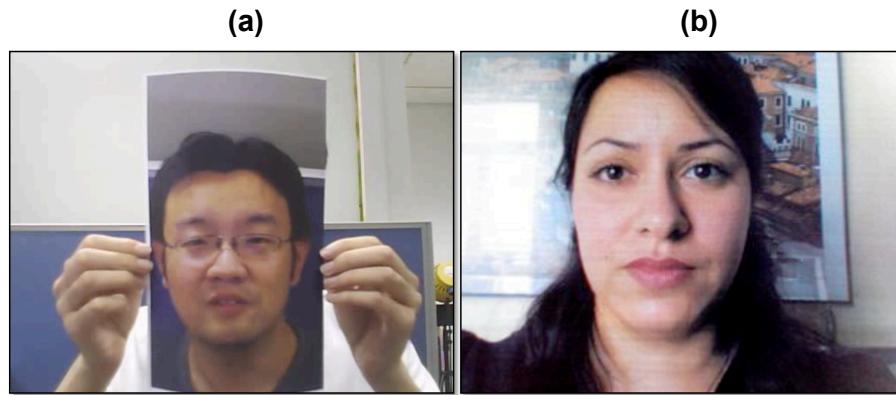


Figure 4.3: Differences in the capture process: (a) CASIA FASD attack (b) Replay Attack Database attack

In the capture bias; the attacks in the CASIA FASD are close-up attacks i.e. the attacker tries to fake only the face region. It is possible to see the borders of the spoofing medium and even the hands of the attacker. The attacks in the Replay Attack Database are scenic i.e. the attacker tries to fake the face and the background at the same time in order to better fake a real access. There is no medium borders and no attackers hands. These differences can be observed in Figure 4.3

Still in the capture bias, we can make another observation. In order to generate a good fake representations of a real access (without any medium borders and attackers hands), the designers of the Replay Attack Database, in general, approximate to much the spoofing medium to the camera. It turns out that the size of the faces in the attacks are generally bigger than in the real accesses. Figure 4.4 shows some examples of that observation.

In order to see if that observation is significative in the whole database, we can run the intra-test protocol using, as a feature, only the area of the face bounding box. Table 4.3 shows the performance of this trick countermeasure.

Table 4.3: *HTER(%)* of the trick countermeasure using only the area of the face bounding box applying the intra-test ($D_1 = D_2$) protocol.

Tune D_1	Test D_2	HTER(%)	
		dev	test
Replay	Replay	24.22	19.63
CASIA	CASIA	51.13	53.09

It can be observed that for the Replay Attack Database the performance in the development and in the test set is far from a random behavior. This experiment confirm the bias observed in this database. It is not possible to observe the same shortcoming in the CASIA FASD.

In the attack bias; the CASIA FASD have different kind of attacks and different way to execute an attack compared to the Replay Attack Database. Exclusive to the CASIA FASD are

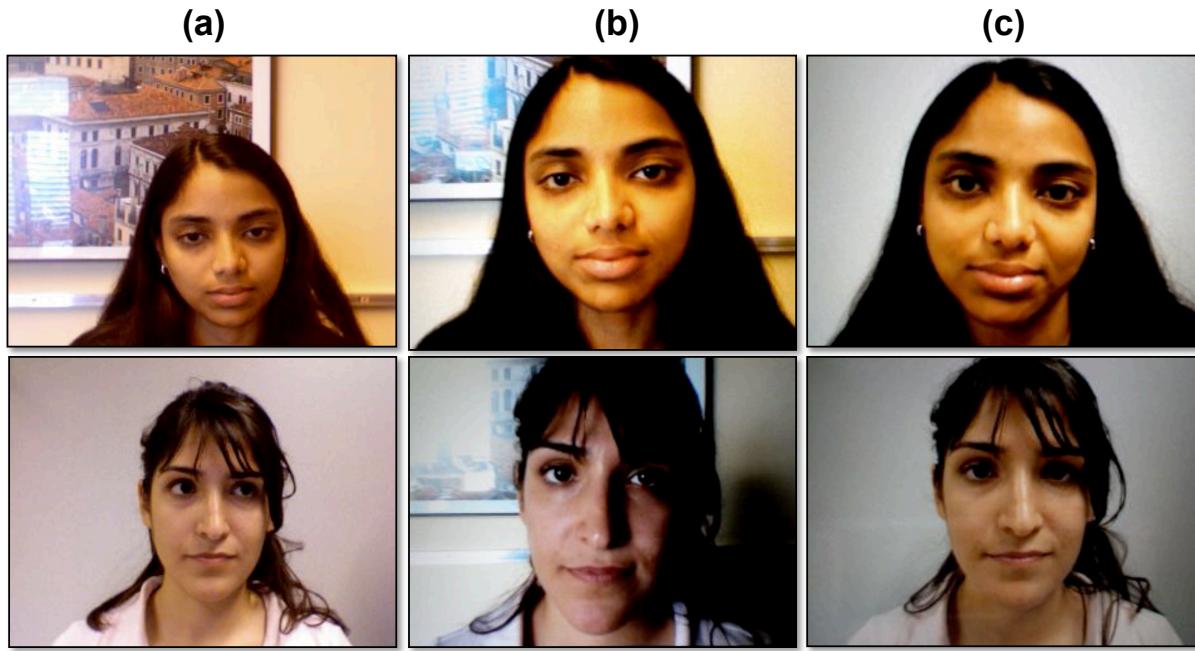


Figure 4.4: Examples of bias in the Replay Attack Database (a) Real access (b),(c) Attempt of attacks

the warped photo and the cut photo attacks which there are no similar attacks in the Replay Attack Database. Exclusive to the Replay Attack Database are the mobile phone attacks. Additionally, the Replay Attack Database has two different support conditions, the fixed and the hand-held. The CASIA FASD has only the hand-held support.

In next section, we will focus if the countermeasures are truly biased to databases or can be tuned to overcome the database bias.

4.5.3 Combination of Multiple Databases

In the previous section we shown that, with the chosen countermeasures, was not possible to get a good performance in both databases at the same time running the inter-test protocol. If we can not get that in tests with databases, what can we say about applying these in a real world scenario? If the databases introduce some bias in the countermeasures due to some particularities of them, we can train each countermeasure with a joint training set combining both databases in order to overcame these biases. Figure 4.6 shows a schematic of this joint training. This is the an intuitive approach to create a more robust countermeasure.

Table 4.4 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 three countermeasures. However, comparing with the intra-test protocol the performance drops drastically. 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.

We can suggest that this strategy is ineffective using these countermeasures. Additionally,

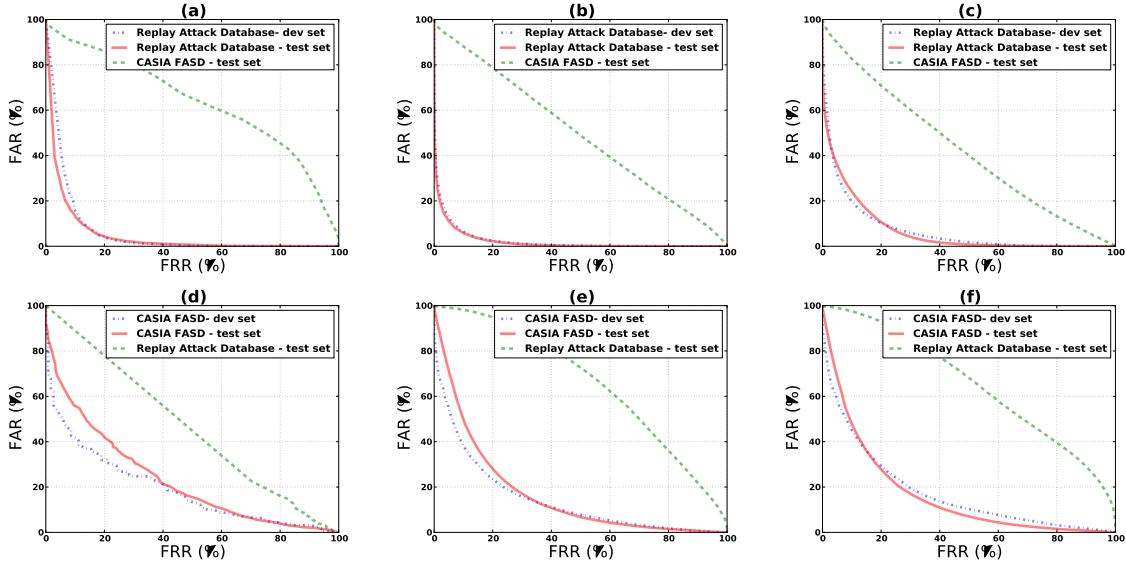


Figure 4.5: ROC curves of each countermeasure using the intra-test and the inter-test protocol. (a) Correlation with frame differences countermeasure trained and tuned with the Replay Attack Database (b) $LBP - TOP$ countermeasure trained and tuned with the Replay Attack Database (c) LBP countermeasure 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.

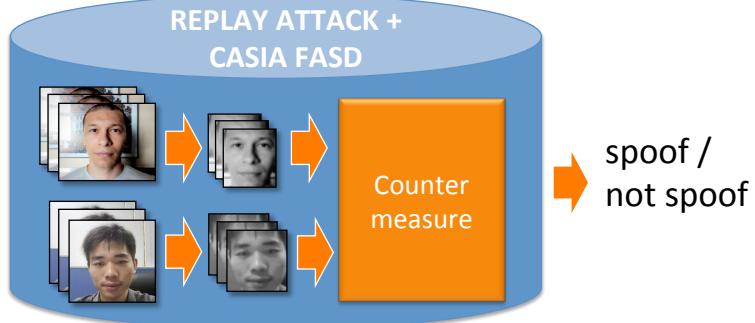


Figure 4.6: Joint training scheme for countermeasures

this strategy has one possible drawback. In face of new kinds of attacks or new databases it is necessary to train and tune all the countermeasures again. And this could be unpleasant.

4.5.4 Score Level Fusion based Framework

In order to improve the performance results in comparison with the intra-test protocol and the inter-test protocol, and to mitigate the bias mentioned in Section 4.2, we introduce a framework based on score level fusion.

This framework consists of training each countermeasure to one specific database; each one

Table 4.4: $HTER(\%)$ 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
Correlation	Replay CASIA	12.18	24.14 43.30
$LBPTOP_{8,8,8,1,1,1}^{u2}$	Replay CASIA	14.29	10.67 42.04
$LBP_{8,1}^{u2}$	Replay CASIA	20.45	19.07 45.92

will generate a score and these scores are fused generating the framework output. The fusion strategy used in this dissertation was a simple sum of normalized scores. Figure 4.7 shows a schema of the Score Level Fusion based Framework. In this Figure, the same countermeasure are trained with two different databases and each one generates a score. These scores are fused generating the final score of the Framework.

Using this framework, when a new countermeasure need to be added, it is possible to "plug it" in the framework. This strategy is similar to an antivirus software. An antivirus is robust against different kind of attacks and they have regular updates in order to become more robust against new threats.

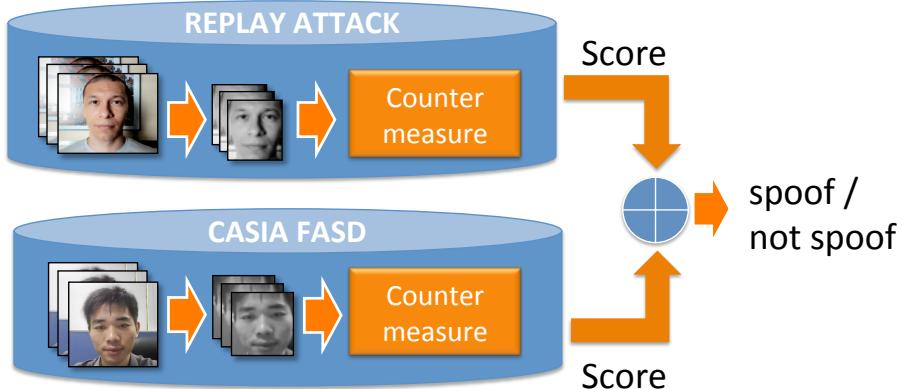


Figure 4.7: Score Level Fusion based Framework schema

To support the assumption of the framework, 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 (Kulcheva & Whitaker 2003) show that the combination of statistically independent classifiers is recommended for a good performance in a score level fusion. In order to evaluate the dependence of classifiers, ten statistics were analyzed. The methodology presented in 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 $Q - statistic$ for two classifiers is defined as follow:

$$Q_{R,C} = \frac{N_{11}N_{00} - N_{01}N_{10}}{N_{11}N_{00} + N_{01}N_{10}} \quad (4.3)$$

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 4.5 shows the statistic dependency using the $Q - statistic$ and the performance in each database trained with the Score Level Fusion based Framework. The analysis is supported with the ROC curves presented in Figure 4.8.

Table 4.5: $Q - statistic$ 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
Correlation	Replay CASIA	0.11 -0.14	13.71	12.39 32.08
$LBPTOP^{u2}_{8,8,8,1,1,1}$	Replay CASIA	0.24 -0.41	23.16	26.04 38.18
$LBP^{u2}_{8,1}$	Replay CASIA	0.38 -0.41	19.69	21.66 47.16

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 (see Table 4.2) and intra-test (see Table 4.1) protocol results. For the inter-test protocol the improvement with the Score Level Fusion based Framework was significative. Comparing with the intra-test protocol the degradation was very low and the countermeasure is able to detect spoofs in both databases with different degrees of success.

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

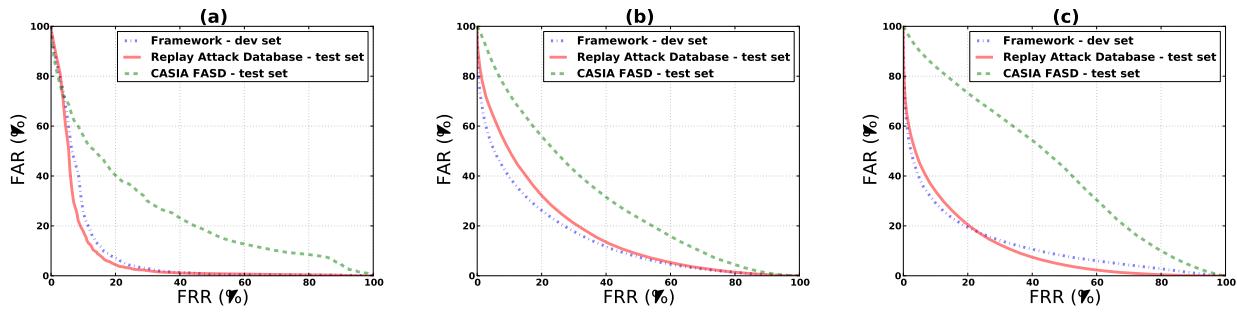


Figure 4.8: ROC curves of each countermeasure trained with the Score Level Fusion based Framework (a) Correlation with frame differences (b) *LBP – TOP* countermeasure (c) *LBP* countermeasure.

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 4.1). The degradation is still high.

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.

4.6 Final Remarks

This chapter compared four countermeasures, very representative according to the state of the art of this research field, using two different test protocols. Using the only two video face antispoofing databases publicly available (Replay Attack Database and CASIA FASD) we introduced the intra-test protocol and the inter-test protocol.

The evaluation of each countermeasure using the intra-test protocol, suggests a good performance and good intra-database generalization power for three countermeasures (Textures with *LBP*, Dynamic textures with *LBP – TOP* and Motion Correlation). The exception was the countermeasure based on eye blinks. Due to some particularities of the databases, this countermeasure was not effective in this protocol and we discarded it. Using the inter-test protocol, the countermeasures accumulates a lot of degradation suggesting a strong bias in the databases. Was highlighted to kinds of database bias, the capture bias and the attack bias.

To overcame these biases we introduce two approaches. The first one, combination of multiple databases, 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. However, it was observed a strong bias to the Replay Attack Database degrading the performance in the CASIA FASD comparing with the intra-test protocol. 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 countermeasure 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. 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 behavior indicated the "improperness" of fusion for these countermeasures.

The Score Level Fusion base Framework can be extended to assume different configurations. For example, it is possible to train different countermeasures with a specific kind of attack. Assuming this configuration, each element of the framework will be specialist to solve one problem (video attacks, mask attacks, printed paper and so on). Additionally it is possible to configure the framework to work with different algorithms. For example, it is possible to fuse the scores of the Motion Correlation with the scores of $LBP - TOP$. It is possible even to provide the score of a face verification as an input for the framework. These different configurations we will be treated in a future work.

Chapter 5

Conclusions

Focusing in antispoofing countermeasures for face authentication, the goal of this masters dissertation is two fold. The first one, we introduce a novel method to detect face spoofing using the spatiotemporal (dynamic texture) extensions of the Local Binary Pattern. The key idea of the approach is to learn and detect the structure and the dynamics of the facial micro-textures that characterise real faces but not fake ones. The second one, is to provide a comparative study of the state of the art countermeasures for face antispoofing. The key contribution of this comparative study is to covers tests in all video face antispoofing databases freely available focusing in the biases that these databases can introduce in the countermeasures.

As emphasized in the beginning, this masters dissertation focuses in a comparative study of antispoofing countermeasures for face authentication.

The Chapter

The intra-test protocol enabled us to measure the performance and evaluate the intra-database generalization of countermeasures. The evaluation of each countermeasure using this protocol, suggests that they are effective to detect spoofs in both databases. Even presenting different performances for different databases, the evaluated countermeasures presented a generalization capability. The only exception was the countermeasures based on eye blinks. With one eye blink, as a liveness check, this countermeasure was easy to deceive it, with a *FAR* higher than 90%. Increasing the number of eye blinks the *FRR* was higher than 90%.

The inter-test protocol enabled us to evaluate the inter-database generalization of countermeasures. Using this protocol, was observed that the evaluated countermeasures accumulates a lot of bias from the databases. It was not possible to detect attacks from one database training the countermeasures with another database. It was observed two kinds of database bias. The first one, called **capture bias**, is a bias related to process of the databases construction. Both databases present different ways to carry out the attacks. The second one, called **attack bias**, is a bias related to the attacks. There are some attacks exclusive to the CASIA FASD and there are some exclusive to the Replay Attack Database.

In order to overcame these biases we introduce two approaches. The first one, combination of multiple databases, combines the train set of each database to train each one of the presented countermeasures. These strategy brought some improvement in performance, in HTER terms, compared with the both protocols, but the database bias still remains. In the second

approach, we introduced the Score Level Fusion based Framework that merges the scores of countermeasures trained with different databases. The results obtained with the Score Level Fusion based Framework suggest that combining two good and not correlated countermeasures leads to significant improvement in performance, in HTER terms, compared with both protocols.

5.1 Contributions

This masters dissertation provided the following contributions:

1. A reproducible research, releasing the all source code
2. A promising countermeasure to detect spoof in a video sequences
3. A method for compare

5.2 Future work

Appendix **A**

Texture Description with Local Binary Patterns

The Local Binary Pattern (LBP) texture descriptor (Ojala et al. 1996), is computed in a pixel level basis using a $n \times n$ kernel, thresholding the surroundings of each pixel with the central pixel value and considering the result as a binary value. The decimal form of the LBP code is expressed as:

$$LBP(x_c, y_c) = \sum_{n=0}^{N-1} f(i_n - i_c)2^n, \quad (\text{A.1})$$

where i_c corresponds to the gray intensity of the center pixel (x_c, y_c) , N is the number of sampling points, i_n is the gray intensity of the n -th surrounding pixel, and $f(x)$ is defined as follows:

$$f(x) = \begin{cases} 0 & \text{if } x < 0 \\ 1 & \text{if } x \geq 0 \end{cases}. \quad (\text{A.2})$$

Later, Ojala et al. (Ojala et al. 2002) extended this operator to support surrounding points and radius of a pixel neighbourhood with different shapes and sizes, enabling handling textures at different scales. Fig. A.1 illustrates the operator calculation and the points distribution in a circular neighbourhood with radius 2, where the pixel values are bilinearly interpolated whenever the sampling point is not in the center of a pixel.

Another important extension proposed by Ojala et al. (Ojala et al. 2002) was the uniform patterns concept ($u2$). A LBP operator is considered uniform if it contains at most two bitwise transitions 0-1 or 1-0 when viewed as a circular bits chain, and according to Ojala et al. (Ojala et al. 2002), nearly 90 percent of LBP operators observed in images are uniform. In spacial terms, uniform patterns represent some patterns of a texture: spot, flat, area, edge and corner. With an 8-bit representation, there are 58 patterns with at most two bitwise transitions. Fig. A.2, extracted from (?), describes all possible uniform patterns with 8 neighbours.

Ahonen et al. (see (?) and (Ahonen, Hadid & Pietikainen 2006)) adopt the following notation for the LBP operator: $LBP_{P,R}^{u2}$, where the subscript represents the neighbourhood configuration with P sampling points on a circle of radius R , and the superscript $u2$ stands for using only uniform patterns and labelling all non-uniform patterns with a single label.

Recently, LBP have been applied to represent face images, yielding in promising results (?). The face description using LBP consists in computing a histogram of LBP operators, and

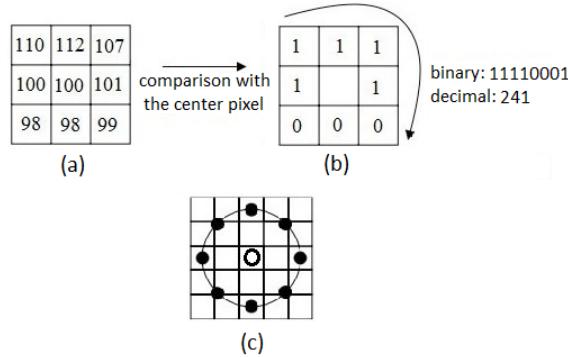


Figure A.1: LBP operator. (a) and (b) The basic LBP operator, where the neighbourhood of each pixel is thresholded and a binary number is obtained. (c) A circular neighbourhood example (with 8 neighbour points and radius 2). The pixel values are bilinearly interpolated whenever the sampling point is not in the center of a pixel.

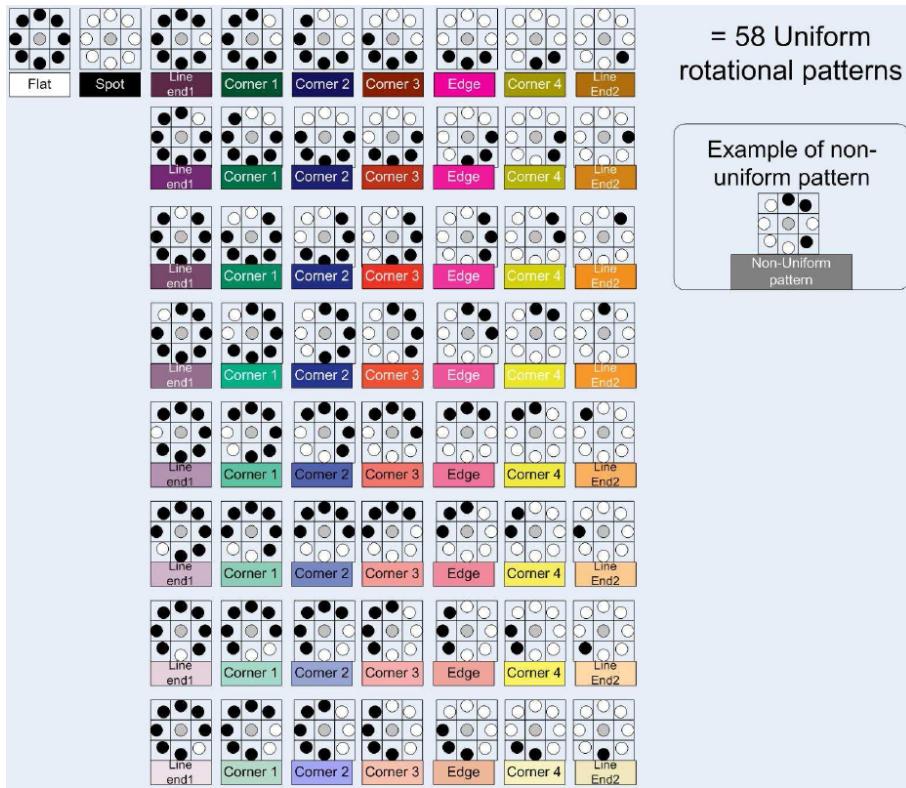


Figure A.2: All uniform patterns for LBP with 8 neighbours (?).

good results have been achieved using a configuration with 8 circular surrounding pixels and radius 2 (see (?), (Ahonen et al. 2006) and (Rodriguez & Marcel 2006)). The LBP histogram-based approach for face description also takes advantage of the LBP operator extensions. As an example, when uniform patterns are used, the number of histogram bins is greatly reduced, since all non-uniform patterns are grouped into a single bin.

A relevant modification in the original LBP operator for face representation is the idea of splitting the face image in small blocks (which can be overlapped or not) and computing the LBP histogram for each block individually, thereby retaining spatial information. So, the face image is described in three different levels: a pixel level, with the calculation of each operator individually; the regional level, with the calculation of histograms for each block; and a global level, with the concatenation of all block histograms (Rodriguez & Marcel 2006). Fig. ?? shows all three levels of face description.

Appendix **B**

Related Publications

Bibliografia

- Ahonen, T., Hadid, A. & Pietikainen, M. (2006). Face Description with Local Binary Patterns: Application to Face Recognition, *IEEE Transactions on Pattern Analysis and Machine Intelligence* **28**: 2037–2041.
- URL:** <http://dx.doi.org/10.1109/TPAMI.2006.244>
- Anjos, A. & Marcel, S. (2011). Counter-measures to photo attacks in face recognition: a public database and a baseline, *International Joint Conference on Biometrics 2011*.
- Bai, J., Ng, T.-T., Gao, X. & Shi, Y.-Q. (2010). Is physics-based liveness detection truly possible with a single image?, *Circuits and Systems (ISCAS), Proceedings of 2010 IEEE International Symposium on*, IEEE, pp. 3425–3428.
- Chakka, M., Anjos, A., Marcel, S. & Tronci, R. (2011). Competition on counter measures to 2-d facial spoofing attacks, *Proceedings of IAPR IEEE International Joint Conference on Biometrics (IJCB), Washington DC, USA*.
- Chetty, G. & Wagner, M. (2004). Liveness verification in audio-video speaker authentication, *Proceeding of International Conference on Spoken Language Processing ICSLP*, Vol. 4, pp. 2509–2512.
- Chingovska, I., Anjos, A. & Marcel, S. (2012). On the effectiveness of local binary patterns in face anti-spoofing, *IEEE BIOSIG 2012*.
- Chingovska, I. et al. (2013). The 2nd competition on counter measures to 2d facial spoofing attacks, *International Conference on Biometrics*.
- da Silva Pinto, A., Pedrini, H., Schwartz, W. & Rocha, A. (n.d.). Video-based face spoofing detection through visual rhythm analysis.
- de Freitas Pereira, T., Anjos, A., De Martino, J. M. & Marcel, S. (2013). Can face anti-spoofing countermeasures work in a real world scenario?
- Duc, N. M., M. B. Q. (2009). Your face is not your password, *Black Hat conference*.
- Eveno, N. & Besacier, L. (2005). A speaker independent” liveness” test for audio-visual biometrics, *Ninth European Conference on Speech Communication and Technology*.

- Flynn, P., Jain, A. & Ross, A. (2008). *Handbook of biometrics*, Springer.
- Froba, B. & Ernst, A. (2004). Face detection with the modified census transform, *Automatic Face and Gesture Recognition, 2004. Proceedings. Sixth IEEE International Conference on*, IEEE, pp. 91–96.
- Galbally, J., Ortiz-Lopez, J., Fierrez, J. & Ortega-Garcia, J. (2012). Iris liveness detection based on quality related features, *Biometrics (ICB), 2012 5th IAPR International Conference on*, pp. 271–276.
- Hill, C. J. (2001). Risk of masquerade arising from the storage of biometrics, *Bachelor of Science thesis, The Department of Computer Science, Australian National University*.
- Inen, M., Pietikäinen, M., Hadid, A., Zhao, G. & Ahonen, T. (2011). *Computer Vision Using Local Binary Patterns*, Vol. 40, Springer Verlag.
- Kollreider, K., Fronthaler, H. & Bigun, J. (2009). Non-intrusive liveness detection by face images, *Image and Vision Computing* **27**(3): 233–244.
- Komulainen, J., Anjos, A., Marcel, S., Hadid, A. & Pietikäinen, M. (2013). Complementary countermeasures for detecting scenic face spoofing attacks.
- Komulainen, J., H. A. P. M. (2012). Face spoofing detection using dynamic texture, *International Workshop on Computer Vision With Local Binary Pattern Variants - ACCV*.
- Kuncheva, L. I. & Whitaker, C. J. (2003). Measures of diversity in classifier ensembles and their relationship with the ensemble accuracy, *Mach. Learn.* **51**(2): 181–207.
URL: <http://dx.doi.org/10.1023/A:1022859003006>
- Leyden, J. (2002). Gummi bears defeat fingerprint sensors, *The Register* **16**.
- Li, J., Wang, Y., Tan, T. & Jain, A. (2004). Live face detection based on the analysis of fourier spectra, *Biometric Technology for Human Identification* **5404**: 296–303.
- Li, S. & Jain, A. (2011). *Handbook of face recognition*, Springer.
- Maatta and, J., Hadid, A. & Pietikaandinen, M. (2012). Face spoofing detection from single images using texture and local shape analysis, *Biometrics, IET* **1**(1): 3 –10.
- Martinez-Diaz, M., Fierrez-Aguilar, J., Alonso-Fernandez, F., Ortega-Garcia, J. & Siguenza, J. (2006). Hill-climbing and brute-force attacks on biometric systems: A case study in match-on-card fingerprint verification, *Carnahan Conferences Security Technology, Proceedings 2006 40th Annual IEEE International*, pp. 151–159.
- Matsumoto, T., Matsumoto, H., Yamada, K. & Hoshino, S. (2002). Impact of artificial gummy fingers on fingerprint systems, *Proceedings of SPIE*, Vol. 4677, pp. 275–289.
- Ojala, T., Pietikäinen, M. & Harwood, D. (1996). A comparative study of texture measures with classification based on featured distributions, *Pattern recognition* **29**(1): 51–59.

- Ojala, T., Pietikainen, M. & Maenpaa, T. (2002). Multiresolution gray-scale and rotation invariant texture classification with local binary patterns, *Pattern Analysis and Machine Intelligence, IEEE Transactions on* **24**(7): 971–987.
- Pan, G., Sun, L., Wu, Z. & Lao, S. (2007). Eyeblink-based anti-spoofing in face recognition from a generic webcam, *Computer Vision, 2007. ICCV 2007. IEEE 11th International Conference on Computer Vision*, IEEE, pp. 1–8.
- Pereira, T. d. F., Anjos, A., De Martino, J. M. & Marcel, S. (2012). Lbp-top based countermeasure against facial spoofing attacks.
- Rodriguez, Y. & Marcel, S. (2006). Face Authentication Using Adapted Local Binary Pattern Histograms, *European Conference on Computer Vision (ECCV)*, pp. 321–332.
- Tan, X., Li, Y., Liu, J. & Jiang, L. (2010). Face liveness detection from a single image with sparse low rank bilinear discriminative model, *Computer Vision–ECCV 2010* pp. 504–517.
- Wei, Z., Qiu, X., Sun, Z. & Tan, T. (2008). Counterfeit iris detection based on texture analysis, *Pattern Recognition, 2008. ICPR 2008. 19th International Conference on*, pp. 1–4.
- Xiao, Q. (2005). Security issues in biometric authentication, *Information Assurance Workshop, 2005. IAW'05. Proceedings from the Sixth Annual IEEE SMC*, IEEE, pp. 8–13.
- Zhang, H., Low, C., Smoliar, S. & Wu, J. (1995). Video parsing, retrieval and browsing: an integrated and content-based solution, *Proceedings of the third ACM international conference on Multimedia*, ACM, pp. 15–24.
- Zhang, Z., Yan, J., Liu, S., Lei, Z., Yi, D. & Li, S. (2012). A face antispoofing database with diverse attacks, *Biometrics (ICB), 2012 5th IAPR International Conference on Biometrics*, IEEE, pp. 26–31.
- Zhao, G. & Pietikainen, M. (2007). Dynamic texture recognition using local binary patterns with an application to facial expressions, *Pattern Analysis and Machine Intelligence, IEEE Transactions on* **29**(6): 915 –928.
- Zhu, Q., Chatlani, N. & Soraghan, J. (2012). 1-d local binary patterns based vad used in hmm-based improved speech recognition, *Signal Processing Conference (EUSIPCO), 2012 Proceedings of the 20th European*, pp. 1633–1637.