

Tiago de Freitas Pereira

X 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

3.1	Selection of face regions	xxxix
3.2	Block diagram of the countermeasure based on LBP	xxxiii
3.3	Block diagram of the DoG countermeasure	xxxiv
3.4	Ataques efetuados com fotografias na base de dados NUAA	xxxv
3.5	Some frames of real access and spoofing attempts (courtesy of IVANA).	xxxvi
3.6	Example images of real accesses and the corresponding spoofing attempts (courtesy of ZHANG)	xxxvii

List of Tables

2.1	Comparison of the most used biometric traits	xxvi
3.1	Performance in $HTER(\%)$ terms of the LBP countermeasure in three face spoofing databases.	xxxiii
3.2	Number of videos in each subset. Numbers displayed as sums indicate the amount of hand-based and fixed support attack available in each subset.	xxxvii

Acronyms

Contents

1	Introduction	xxiii
1.1	Scope and Contributions	xxiii
1.2	Organization of the Thesis	xxiii
2	Biometrics	xxv
2.1	Introduction to Biometric Systems	xxv
2.2	Attacks in Biometric Systems	xxvi
2.2.1	Replay attack	xxvi
2.2.2	Biometric reference attack	xxvii
2.2.3	Man-in-the-middle attack	xxvii
2.2.4	Ataque de Spoofing	xxvii
3	Spoofing Attacks	xxix
3.1	Spoofing Attacks in Biometrics	xxix
3.1.1	Fingerprint	xxix
3.1.2	Speaker	xxix
3.1.3	Iris	xxx
3.2	Spoofing Attacks in Face Recognition	xxx
3.2.1	Presence of vitality (liveness detection)	xxx
3.2.2	Scene	xxxii
3.2.3	Differences in image quality assessment	xxxiii
3.3	Face Spoofing Databases	xxxiv
3.3.1	NUAA	xxxv
3.3.2	Replay-Attack Database	xxxv
3.3.3	CASIA Face Anti-Spoofing Database	xxxvi
4	Developed Countermeasures	xxxix
5	The Comparative Study	xli
5.1	Databases Permeability	xli
5.2	Evaluation Protocol	xli
5.2.1	Intra Database Test Protocol	xli

5.2.2	Inter Database Test Protocol	xli
5.2.3	Evaluation Metrics	xli
6	Experiments and Results	xliii
7	Conclusion	xlvi
8	Future Work	xlvi
A	Related Publications	xlix
	References	1

Chapter 1

Introduction

1.1 Scope and Contributions

1.2 Organization of the Thesis

Biometrics

Some text.

2.1 Introduction to Biometric Systems

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 (?). The use of biometrics as authentication factor has some advantages. Naturally, is not possible to forget or transfer a biometric trait and it hardly disappears (perhaps in case of a seriously accidents). Biometrics has some disadvantages. As an example, our voice vary drastically when we get sick or when we are under stress. Unfortunately our facial trait change when we get old. It is important to remark that authentication based on biometrics is probabilistic, which means that errors can happen (????).

To use a biometric trait in a biometric system, the candidate must satisfy the following requirements.

- Universality (every person must have it);
- Uniqueness (must distinguish people);
- Stability (must be stable along the time);
- Coletability (must be measure);
- Performance;
- Acceptance;
- Circunvention (low risk of frauds).

Table 2.1 shows a comparative between the most used biometric traits (?). It can be observed that none of the presented biometric traits fulfil all the listed requirements and the selection of a trait depends on the application (?????) (?).

Table 2.1: Comparison of the most used biometric traits (?)

Biometric trait	Universality	Uniqueness	Stability	Coletability	Performance	Acceptance	Circunvention
Face	High	Low	Medium	High	Low	High	Low
Fingerprint	Medium	High	High	Medium	High	Medium	Medium
Hand geometry	Medium	Medium	Medium	High	Medium	Medium	Medium
Palm vein	Medium	Medium	Medium	Medium	Medium	Medium	High
Iris	High	High	High	Medium	High	Low	High
Signature	Low	Low	Low	High	Low	High	Low
Voice	Medium	Low	Low	Medium	Low	High	Low

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

FIGURE X

Firstly the biometric trait is captured using some sensor. Secondly the captured biometric trait is processed in order to extract the biometric features. When it is in an enrolment procedure, these features will generate a biometric reference, and it will be stored in a database. In an authentication procedure, these features will be used in a comparison with the stored biometric reference. It is possible to observe in the Figure Y, attacks can be done in any point of the architecture (?). The next subsections will be discussed about each one of the possible point of attacks and how to mitigate it.

FIGURE Y

2.2 Attacks in Biometric Systems

2.2.1 Replay attack

The replay attack is performed by injecting a biometric data previously sent, of the target identity, in order to have a non authorised access. This data can be obtained sniffing the biometric software in one of the points of the Figure X. To mitigate these kind of attacks, the biometric system should ensure that the provided data was not injected artificially (?). The most popular way of protect this kind of attack is to associate a timestamp to the data. As it is improbable to have the exactly the same biometric data in different times, this method is effective.

2.2.2 Biometric reference attack

The attack in the biometric reference is performed where the biometrics are stored. This kind of attacks include actions such as the inclusion, removing, modifying and steal a biometric template. Among this actions, the possibility to steal a biometric reference is the most dangerous treat, since it is possible to work in a reverse engineering processes to regenerate the biometric trait. (?) shown that is possible to generate synthetic fingerprints applying reverse engineering in a biometric references based on minutia. With this fake fingers was possible to deceive a fingerprint authentication system. To mitigate the risk of this kind of attack is recommended to encrypt the biometric reference and increase the policy to access the biometric references.

2.2.3 Man-in-the-middle attack

The man i

2.2.4 Ataque de Spoofing

The spoofing attack in biometric system is a direct attack to the biometric sensor; a forged biometric is presented to the biometric sensor. The goal is to pretend to be someone else in order to get some forbidden privileges.

Best practices in security recommends to use cryptography or the creation of security channels in order to mitigate most of the problems aforementioned. But in case of a spoofing attack, the target is the biometric sensor, and in the chart presented in Figure X, is not possible to apply any of the security tools to prevent attacks, becoming the most fragile point of attack. This kind of attacks is the main point of this thesis.

Spoofing Attacks

As aforementioned, spoofing attacks in biometrics are direct attacks to the biometric sensor. Spoofing techniques vary from different biometrics. This chapter discusses the spoofing attacks in different biometric traits focusing in face recognition.

3.1 Spoofing Attacks in Biometrics

It is always possible to create a spoofing attack to a biometric system. This sections discusses the occurrences of spoofing attacks in different biometric systems.

3.1.1 Fingerprint

In fingerprints based systems the attacker can forge a fingerprint with different materials (gummy, silicone, etc) in order spoof the system as can be observed. (?) and (?) discusses how to generate fake fingerprints using materials that is possible easily to find in a supermarket. Figure X shows how easy is to create a mold from a live finger and to reproduce its fingerprint with gummy. This fake fingers can be used to spoof a fingerprint biometric systems.

Figure X.

Recently in Brazil (2013), was reported that doctors in Sao 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 clock¹.

A more sophisticated attack were discussed in (?). This paper uses a hill climbing procedure to optimise the position and the orientation of the minutia in a minutia based fingerprint verification system. This optimized minutia can be used to generate a fake finger.

FALAR DOS SENSORES.

3.1.2 Speaker

For the 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. (?) e (?) address the problem

¹<http://www.foxnews.com/us/2013/03/13/brazilian-doctors-use-fake-silicone-fingers-to-defraud-hospital-punch-in-clock/>

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.

PAPER analyses the speech signal itself applying the 1D-LBP (Local Binary Pattern) followed by a SVM (Support Vector Machines) in order to detect spoofs.

3.1.3 Iris

(?), (?) and (?) are works addressing spoofing attacks in iris biometric system.

3.2 Spoofing Attacks in Face Recognition

Because of its natural and non-intrusive interaction, identity verification and recognition using facial information are among the most active and challenging areas in computer vision research. Despite the significant progress of face recognition technology in the recent decades, wide range of viewpoints, ageing of subjects and complex outdoor lighting are still research challenges. Advances in the area were extensively reported in (?) and (?).

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 (?, ?, ?, ?) and contests addressing the problem. Two contest were organized in the last two years. The first one was organized under IJCB 2011(International Joint Conference on Biometrics)(?) 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).

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. In order to mitigate the vulnerability of face authentication systems, effective countermeasures against face spoofing have to be deployed.

Recently, the media has documented some situations of attacks in deployed face recognition systems. Using simple photographs, a research group from University of Hanoi showed how easy is to spoof the face authentication systems deployed in Lenovo, Asus and Toshiba Laptops (?). 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 system². As a consequence, a eye blinking detection has been introduced in the most recent version of the Android OS.

²<http://www.itproportal.com/2011/11/14/ice-cream-sandwich-facial-recognition-cracked/>

The countermeasures against spoofing attempts in face recognition can be macro classified between the countermeasures that depends of user collaboration and the countermeasures that do not depends of user collaboration. The countermeasure that depends of 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 expressions³. This kind of strategy can be fun in the beginning but in some situations this can be embarrassing. The countermeasures that do not depend of user collaboration, try to solve this issue analysing the video 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.

3.2.1 Presence of vitality (liveness detection)

Presence of vitality or liveness detection, consists of search for features that only live faces can possess. The eye blinking is an activity that humans do constantly. A regular human 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, doesn't matter in what condition the person is, in some point the eye blink will occur. Following that fact, (?) 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 download⁴, shown an accuracy of 95.7% .

Supported by the hypothesis that live faces present uncorrelated motion patters in some parts of the face compared to the attacks, (?) 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 3.1.

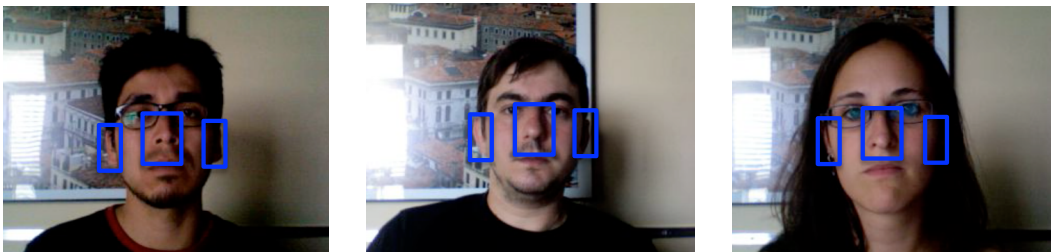


Figure 3.1: Selection of face regions of the algorithm (?)

The strategy of the countermeasure can be summarized as follows:

³<http://www.bbc.co.uk/news/technology-22790221>

⁴http://www.cs.zju.edu.cn/gpan/database/db_blink.html

1. Detect the face region;
2. Delimitate the region of the face center and the ears (Figure 3.1);
3. Determine if the face region is moving more horizontally or more vertically analysing the optical flow velocities;
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 database⁵ and the attacks were generated with printed photographs from the same database. With this database, which was not made public, an $EER = 0.5\%$ was achieved.

EULERIAN.

3.2.2 Scene

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

The countermeasure proposed in (?)⁶ 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:

$$M_D = \frac{1}{S_D} \sum_{(x,y) \in D} |I_t(D) - I_{t-1}(D)|, \quad (3.1)$$

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

To input the motion coefficient into a classifier, 5 quantities are extracted for every window of 20 frames. The quantities are: the minimum of M_D in that time window, the maximum, the average, the standard deviation and the ratio R between the spectral sum for all non-DC components and DC component itself taken as base the N -point Fourier transform of the signal (see Equation 3.2). These 5 quantities are fed into a Multi-layer Perceptron (MLP) classifier with 5 neurons in the hidden layer which is trained to detect spoofing attacks. This countermeasure was evaluated using the photograph attacks subset of the Replay Attack Database(?) and achived an $HTER = 9\%$.

$$R = \frac{\sum_{i=1}^N |FFT_i|}{|FFT_0|} \quad (3.2)$$

⁵<http://www.ee.surrey.ac.uk/CVSSP/xm2vtsdb/>

⁶<http://pypi.python.org/pypi/antispoofing.motion/>

3.2.3 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, (?) and (?), explored the *LBP* (Local Binary Patterns) texture descriptor analysing single frames. In this countermeasure the detected faces (see Figure 3.2) 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.

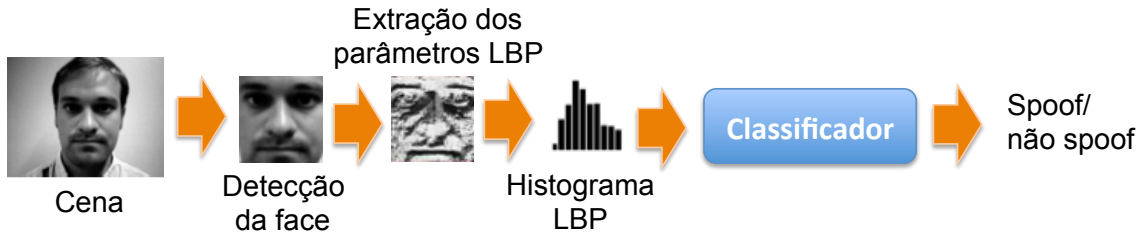


Figure 3.2: Block diagram of the countermeasure based on LBP

The Table 3.1 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. It can be observed a satisfactory performance in the three databases (between $\sim 15\%$ and $\sim 20\%$). However, comparing the performance in the development set (used to tune the hyperparameters) and in the test set of the NUAA database suggest a low generalisation capability.

Table 3.1: 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, (?) propose a countermeasure based on Difference of Gaussians filters (DoG).

As can be observed in the block diagram in Figure 3.3, 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 $\sigma_2 = 2$.

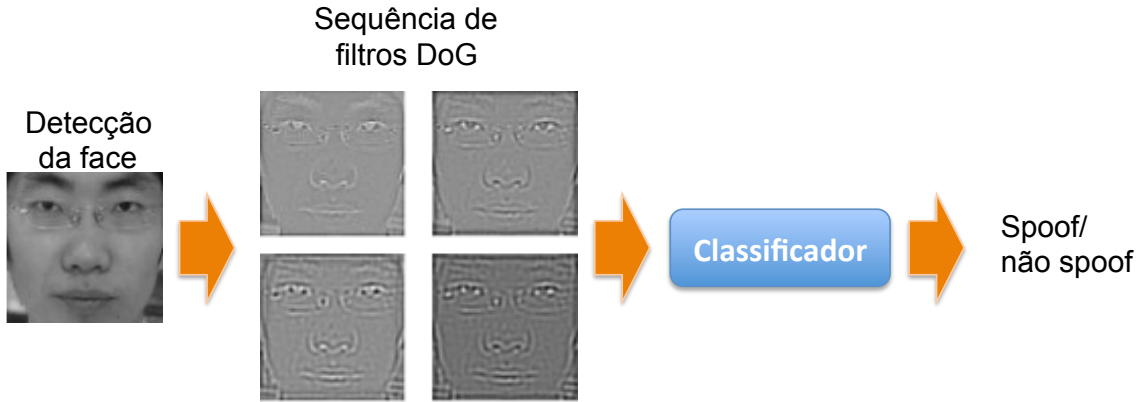


Figure 3.3: Block diagram of the DoG countermeasure

After of the sequence of filters, the images are geometric normalised to 128×128 pixels and these data fed into a *SVM* classifier. Evaluated using a the CASIA FASD, the countermeasure show an *EER* of 17%.

Li et al. (?) 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, an accuracy of 100% was achieved.

In order to detect noise patterns in spoofing attacks, (?) 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(?). This technique, generates one image with a combination of all frames. These steps end the preprocessing steps.

A texture description using Gray Level Cooccurrence matrix (GLMC) was applied in the Visual Rhythm image. With the cooccurrence 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.

3.3 Face Spoofing Databases

In this section, we give an overview of the two largest and most challenging face spoofing databases, Replay-Attack Database (?) and the CASIA Face Anti-Spoofing Database (?),

consisting of real access attempts and several fake face attacks of different natures under varying conditions. Instead of still images, both datasets contain short video recordings which makes them suitable for evaluating countermeasures that exploit also temporal information.

3.3.1 NUAA

The NUAA face spoofing database⁷ 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 3.4 has some examples of this database.



Figure 3.4: Ataques efetuados com fotografias na base de dados NUAA

3.3.2 Replay-Attack Database

The Replay-Attack Database⁴ (?) 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.

⁷<http://parnec.nuaa.edu.cn/xtan/data/NuaaImposterdb.html>

The support conditions include:

1. **hand-based**: the attacker holds the attack media using his own hands;
2. **fixed**: the attacker sets the attack device in a fixed support so it does not move during the spoofing attempt.

Fig. 3.5 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 3.5: Some frames of real access and spoofing attempts (courtesy of IVANA).

The Replay-Attack database provides a protocol for objectively evaluating a given countermeasure. Such protocol defines three non-overlapping partitions for training, development 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 3.3).

$$HTER = \frac{FAR(\tau, D) + FRR(\tau, D)}{2}, \quad (3.3)$$

where τ is a 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.

3.3.3 CASIA Face Anti-Spoofing Database

The CASIA Face Anti-Spoofing Database⁵ (?) contains 50 real clients and the corresponding fake faces are captured with high quality from the original ones. The variety is achieved by introducing three imaging qualities (low, normal and high) and three fake face attacks which include warped photo, cut photo (eyeblick) and video attacks. Examples from the database can be seen in Fig. 3.6. 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

Table 3.2: 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

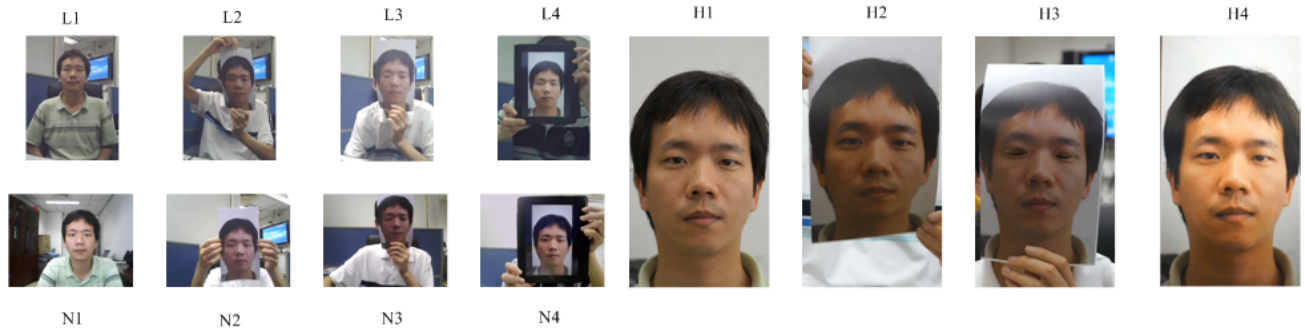


Figure 3.6: Example images of real accesses and the corresponding spoofing attempts (courtesy of ZHANG)

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.* (?).

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.

Chapter 4

Developed Countermeasures

The Comparative Study

5.1 Databases Permeability

5.2 Evaluation Protocol

5.2.1 Intra Database Test Protocol

5.2.2 Inter Database Test Protocol

5.2.3 Evaluation Metrics

Chapter 6

Experiments and Results

Chapter 7

Conclusion

Chapter 8

Future Work

Appendix A

Related Publications

Bibliografia