Protection of 2D face identification systems against spoofing attacks

Taiamiti Edmunds

UNIVERSITÉ GRENOBLE ALPES

THÈSE

pour obtenir le grade de

DOCTEUR DE LA COMMUNAUTÉ UNIVERSITÉ GRENOBLE ALPES

Spécialité : **Signal, Image, Parole, Télécoms**

Arrêté ministériel : 7 août 2006

Présentée par
Taiamiti EDMUNDS

Thèse dirigée par **Alice CAPLIER**

préparée au sein du
Grenoble Images Parole Automatique (GIPSA-LAB) dans **l'Ecole Doctorale d'Electronique, Electrotechnique,Automatique et Traitement du Signal (EEATS)**

Protection of 2D face identification systems against spoofing attacks

Thèse soutenue publiquement le **23 janvier 2017**,
devant le jury composé de:

Catherine ACHARD
MCF ISIR, Rapporteur
Jean-Luc DUGELAY
PR EURECOM, Rapporteur
Franck DAVOINE
MCF HEUDIASYC, Examinateur
Pierre-Yves COULON
PR Grenoble-INP, Examinateur, Président du jury
Alice CAPLIER
PR Grenoble-INP, Directrice de thèse

THÈSE

pour obtenir le grade de

DOCTEUR DE L'UNIVERSITÉ DE GRENOBLE ALPES

Spécialité : **Signal, Image, Parole, Télécoms**

Arrêté ministériel : 7 août 2006

Présentée par
Taiamiti EDMUNDS

Thèse dirigée par **Alice CAPLIER**

préparée au sein du
Grenoble Images Parole Automatique (GIPSA-LAB)
dans **l'Ecole Doctorale d'Electronique, Electrotechnique,Automatique et Traitement du Signal (EEATS)**

Protection of 2D face identification systems against spoofing attacks

Thèse soutenue publiquement le **23 janvier 2017**,
devant le jury composé de:

Catherine ACHARD
MCF ISIR, Rapporteur
Jean-Luc DUGELAY
PR EURECOM, Rapporteur
Franck DAVOINE
MCF HEUDIASYC, Examinateur
Pierre-Yves COULON
PR Grenoble-INP, Examinateur, Président du jury
Alice CAPLIER
PR Grenoble-INP, Directrice de thèse

Remerciements

Je souhaite remercier tout d'abord ma directrice de thèse Alice Caplier pour son soutien et son encadrement tout au long de ce travail. Merci Alice pour ta patience et tes encouragements. J'ai vraiment apprécié la grande liberté et l'autonomie que tu m'as accordées pour effectuer mes recherches.

Je tiens à remercier également les membres de mon jury de thèse pour avoir accepté d'examiner mon travail. Merci à Catherine Achard et Jean-Luc Dugelay pour leurs remarques constructives qui m'ont permis d'améliorer la qualité du manuscrit. Je tiens à remercier aussi Franck Davoine et Pierre-Yves Coulon pour l'intérêt qu'ils ont porté à mes travaux en qualité d'examinateurs.

Cette thèse n'aurait pas pu être réussie sans l'aide de tous les collègues et amis du Gipsa-lab. A commencer par les ex deuxième et troisième années au moment de mon arrivée en thèse, notamment Céline et Cindy qui m'ont aidé pour les démarches administratives avant même mon arrivé au laboratoire! J'ai ainsi bénéficié d'une intégration rapide au sein des doctorants du DIS, avec également l'aide d'autres anciens SICOMs: Lucas et Alexis! Evidement je suis obligé de mentionner aussi les autres membres du bureau D1146 (Arnaud, Raluca, Pascal et Tuan) ainsi que Manu, Tim(s), Florian et Guillaume. Egalement, je dois remercier Jérémy et Pedro qui ont été de chouettes et brillants co-bureaux!! Ils ont été d'une grande aide dans mes recherches avec aussi Lucas le dieu de l'optimisation ;). Puis, merci à tout le groupe WhatsApp GipsaKing sans qui les sorties bar n'auraient pas été aussi géniales :), avec une spéciale dédicace à Tim2 pour les meilleures soirées du lundi soir :) mais aussi à Miguel, Victor, Paolo, Quentin et Marielle qui ont redonnés vie aux "after-works". Merci aussi au Gipsa-doc pour l'organisation des évènements Gipsa qui a transformé le labo un lieu de travail vivant et sympa.

Je me dois de mentionner ma nouvelle famille de la BroLoc: Florian, Hadrien, Kevin, Marvin et Yann. Ces années n'auraient pas été aussi joyeuses et folkloriques sans eux, leurs copines et leurs familles.

Enfin, je remercie chaleureusement ma famille pour leur soutient tout au long de mes études. Malgré la distance et les difficultés pour s'avoir au bout du fil, vous serez toujours ceux qui comptent le plus et à qui je dois cette réussite! Merci papa d'être venu assister à ma soutenance!

Contents

List of Figures

List of Tables

Introduction

Biometric systems have invaded our daily lives as many applications have now replaced the traditional badges and passwords. As time goes by, new biometric solutions have emerged going from voice and face recognition to fingerprint, iris and veins recognition. In this thesis, we are interested in face biometric solutions which cover a wide range of applications [Parmar14] such as face identification, access control, border control, surveillance and general identity verification. For instance, face recognition technology is used in the Fresno Yosemite International airport to alert the authorities if a known terrorist is recognized by the system. Another example concerns the elimination of duplicates during the voter registration procedure as each voting member is assigned a unique ID number only if no match is found among already registered individuals. Otherwise manual inspection is conducted. The most common use of face biometric solution concerns access control applications where the face replaces the traditional password and login for many use-cases such as banking, room and facility access, computer and smart-phone login. The success of security solutions based on face biometrics is based on the fact that face is a unique biological identity marker (except for twins) which establishes a connection between the identification number and password contrary to conventional security systems. Hence, multiple face biometric commercial systems have conquered the security market [Gorodnichy14] and more than a hundred companies are referenced on the internet including large groups such as NEC [1] or Google [2]. Beyond the competition for improving face recognition performance, new challenges have emerged regarding the security of biometric solutions in general. Major security weaknesses have been unveiled to the general public through real spoofing attempts on breakthrough systems such as airport security scans [3] or commercial laptops [Duc09]. Even spoofing demonstrations are available online on Youtube [4]. The biometric community has addressed these security issues for a decade now with the support of two European projects BEAT [5] and TABULA-RASA [6]. The development of diverse spoofing attacks has been studied in the context of the TABULA-RASA Spoofing Challenge in 2013. Along the way, the construction of public databases and the development of protection measures against spoofing attacks have been encouraged by the the two ICB competitions in 2011 [Chakka11] and 2013 [Chingovska13b]. In this context, multiple anti-spoofing countermeasures have been introduced bringing together different research fields such as computer vision, texture analysis, motion analysis and optic physics.

0.1 Face recognition system's vulnerabilities

A face recognition system and more generally any biometric system comprise two stages: the enrolment and the authentication. The first step enables an unknown user to register on the system's database by following the standard authentication procedure. A template of the user in the form of face images or any other face representations is stored in the database. In the authentication phase, two different matching strategies are implemented depending on the operating instructions of the face recognition system. The identification mode searches the identity in the system's database

[1] http://www.nec.com
[2] http://www.pittpatt.com
[3] http://edition.cnn.com/2010/WORLD/americas/11/04/canada.disguised.passenger/
[4] https://www.youtube.com/watch?v=KSHV23aPm2s
[5] https://www.beat-eu.org/
[6] https://www.tabularasa-euproject.org/

that best matches the input face and a one-versus-all comparison procedure is conducted. The verification mode corresponds to the case where the user claims his identity and the system simply checks that his face matches with his corresponding template using a one-versus-one comparison procedure. The latter case is what we focus on in this thesis and in face anti-spoofing in general. The general architecture of a face biometric system comprises a sensor (RGB camera in our case), a feature extractor module and a matcher in communication with an off-line or on-site database. Ratha et al. [Ratha01a, Ratha01b] have identified eight basic weaknesses in a biometric system as depicted in figure 1.

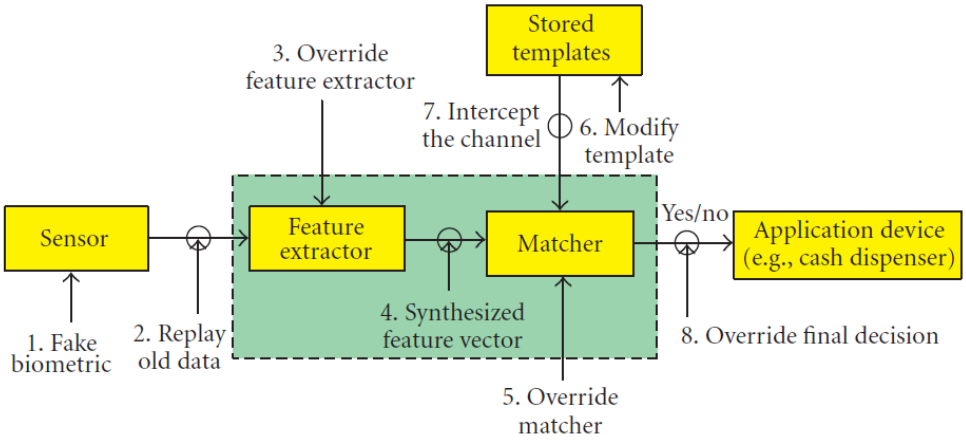


Figure 1: Weaknesses of biometric systems in general according to [Ratha01a].

These points of attacks are grouped into four categories in [Jain08] as follows:

- attacks at the user interface (1): the impostor presents a fake face (photo, video or masks) of a valid user in front of the sensor.
- attacks at the interfaces between modules (2,4,7,8): communication channels between modules are intercepted and tampered to simulate a valid access such as hill climbing attacks [Soutar02, Galbally10, Gomez-Barrero12].
- attacks on the modules (3,5): the behaviour of modules can be altered so they produce a specific response, these attacks are known as Trojan-horse attacks.
- attacks on the template database (6): an impostor can gain access to the system by replacing the template of a valid user in the database with his own template.

The last three types of attacks are out of the scope of this work as they are not specific to biometric systems but rather to any security system. The term spoofing commonly used in the biometric community designates attacks at the sensor by presentation of a fake face. This vulnerability is very specific to biometric systems and is particularly problematic for face biometric solutions as manufacturing a fake face is very easy and low cost compared to other biometrics. Videos or pictures can be easily stolen, either by taking a picture directly without the targeted person's consent or directly on the internet as social networks are gathering all this kind of data. Therefore, someone's face is no longer a secure information and can be hijacked to spoof face recognition systems directly by presenting a stolen photo or video. Not every picture can spoof such a system because it must fulfil certain matching criteria based on the enrolment process. However, by testing a few pictures/videos with different lightning conditions, points of view and backgrounds (easy to do with image processing tools) one can spoof a recognition system quite

easily. In [Duc09], Duc and Minh demonstrate the vulnerability of three face verification systems from commercial laptops against photo-spoofing. This vulnerability has been known for a long time and has drawn a lot of attention this last decade.

0.2 Context: the BIOFENCE project

The evaluation and certification of spoofing resistance are two of the major issues concerning biometrics technologies in their present and future implementations. The ANR BIOFENCE project proposes a systematic study of the spoofing resistance of face, iris and vein patterns biometrics in order to come up with a suitable evaluation methodology for certification.

The first part of BIOFENCE concerns the technical analysis of existing spoofing attacks as well as the creation or development of new fakes we could foresee for face, iris and vein biometrics. If the literature reports several cases of attacks to fingerprint biometric sensors by fake fingers, information regarding face recognition are rarer, and even more for iris and vein modalities. This is naturally due to the fact that biometric sensors using the fingerprint are more prevalent in the world and to some extent it is this type of sensors that have laid the foundations of modern biometrics. The project aims to draw up an exhaustive inventory of spoofing attacks aiming at spoofing these modalities (status of technical and scientific advances, patents, publications) and to design scenarios in order to assess them in terms on ease of implementation, ease of use, impact, etc.

The second objective of this project is to provide a comprehensive and innovative study of countermeasures. From an industrial point of view, a biometric provider wanting to certify a product will aim for the maximum security level, i.e. will rely on the implementation of innovative protection techniques. These countermeasures intend to improve resistance to attacks using spoof of the state of the art but also ideally new fakes that will be proposed during this project. Therefore, based on a review of existing countermeasures or liveness detection techniques, the BIOFENCE project will have to deal with a bunch of problems ranging from relatively explored to unexplored. The solutions can be hardware or software and each solution must then be evaluated in terms of cost, complexity and skills required to implement it.

The third objective is to lean on the Common Criteria (CC) standard to establish the assessment methodology of robustness for face, iris and veins biometric sensors. The adaptation of this norm to face, iris and vein biometrics will ensure reliable and robust evaluation criteria. CC standard is internationally recognized and this work aims to support the set up of an international standard for security evaluation of biometric systems.

A particular attention is paid to the compliance of the developments in terms of ethics and respect of privacy as well as to the societal impacts of the project. The consortium is composed of one industrial (Safran) and multiple research labs to cover all the biometrics technologies as well as ethical aspects as illustrated in figure 2.

In summary, multiple scientific advances and results are expected from this project:

- Better understanding/knowledge and inventory of present and future (anticipated) spoofing attacks to face, iris and venous network biometrics.
- Enhancement of biometric acquisition systems and software processing techniques enabling

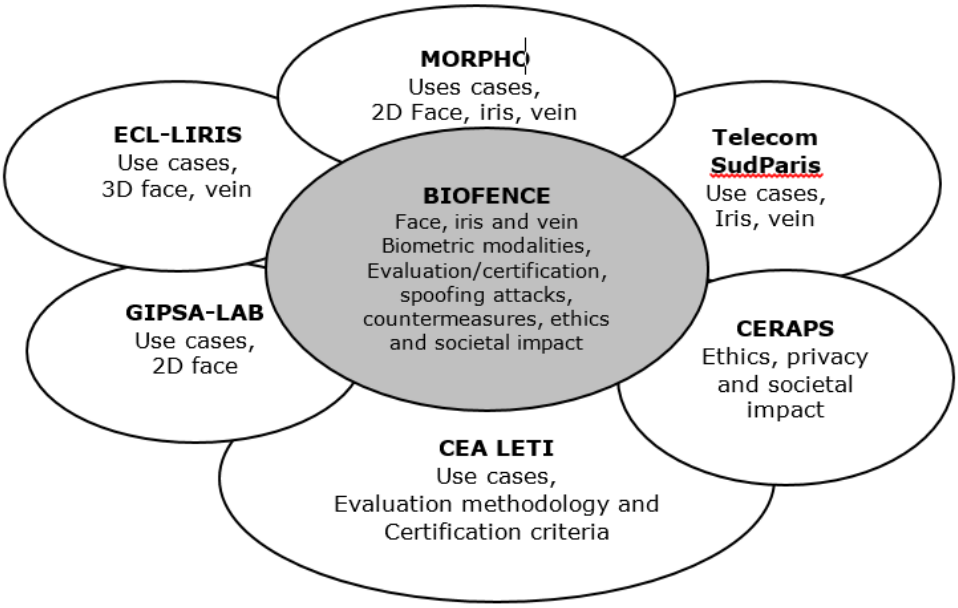


Figure 2: Consortium description: contribution of each partner.

an improved robustness to fakes. This will lead to better security of FIV biometrics to ensure a high level of resistance to attacks for the future products.

- Define a standardized framework for evaluating these systems and propose an adaptation of the Common Criteria norm aiming to set up an international certification standard for security of face, iris and vein pattern biometric products.
- Studying the societal impacts of the project and ensuring privacy.

Our contribution in the BIOFENCE project focuses on 2D face biometric technology and is articulated in three parts. The first challenge is to evaluate the resistance of unprotected face recognition systems against state of the art spoofing techniques and possibly anticipate and develop new attack scenarios. The second objective is to develop new protection measures to improve the resistance of face biometric systems against these treats. Finally, the certification methodology based on the Common Criteria developed for fingerprint technology is applied for 2D face recognition systems security evaluation. Tests are carried out in order to evaluate if the methodology reflects the resistance of protected 2D face recognition systems against spoofing attacks correctly.

0.3 Anti-spoofing challenges

The protection of face biometric systems against spoofing attacks raises a number of challenges as fake faces are easy to forge with minimum equipment. At the moment, numerous anti-spoofing methods exist and most of them work under very restrictive specifications including the authentication protocol, the acquisition conditions, the type of attack considered and even the way the attack is performed. The anti-spoofing problem suffers from many sources of variability that sometimes overcomes the differences between real and fake faces. In a sense, the anti-spoofing problem can be seen as the dual task of the face recognition task. On the one hand, face recognition algorithms strive for a face embedding that expands the interpersonal variabilities (differences between two individuals) while reducing the intra-personal variance generated by illumination, pose and facial

expressions. On the other hand, anti-spoofing countermeasures search for a face representation that reduces the interpersonal variance while increasing the intra-personal variance in a way that limits the variability due to pose and illumination changes but expands the variability due to the forged nature of fake faces. The main difficulty is to determine discriminant cues for all types of attacks and acquisition conditions. At first, liveness based countermeasures have been proposed to solve the photo attack problem. However, as spoofing techniques keep evolving, more complex countermeasures need to be developed to remain one step ahead. The arrival of video attacks and mask attacks have added another level of complexity to the problem at hand. Hence, some countermeasures aim to control certain sources of variability to make them discriminant. For example, head movements (pose variability) or facial expressions such as blinking or smiling can be specified during authentication in a challenge/response procedure to assess liveness. Other methods focus on background or scene context information and are likely to fail against mask attacks. Besides, anti-spoofing performance drops significantly when only the face region is considered.

In this work, we focus on the face region only with controlled illumination conditions (that are consistent between training and testing the system) and under a non-cooperative authentication protocol. We exploit intrinsic differences between natural and unusual characteristics of the face only and study their consistency between different types of attacks and acquisition conditions.

0.4 Main contributions

Several methods combine multiple complementary cues such as motion and texture information from the face region to deal with all sorts of attacks but the contribution of each component is not always clearly specified. Our goal is to isolate and study motion and texture independently to assess their strengths and limitations.

Our first contribution is the review of state of the art methods to draw out the current limitations of protection methods. We limit our focus to software-based methods in this work and survey a large range of methods although only texture-based and motion-based methods are discussed in this document.

Our second contribution is the development of an unified evaluation framework based on the study of the popular LBP descriptor for the evaluation of state of the art texture-based countermeasures. Then, we propose two different approaches to improve the classic LBP descriptor for deriving discriminant features. We propose to extend the LBP operator to embed contrast and color information in the texture characterization. The proposed HSI-LBP color texture descriptor obtains significant improvements compared to state of the art texture-based methods on the ReplayAttack, CASIA and MSU databases.

Our third contribution is the development of a new motion-based countermeasure based on the constrained local neural fields framework of [Baltrusaitis13]. The face is modelled by a deformable shape composed of 68 landmarks. Rigid and non-rigid motions are directly extracted from this convenient face tracking framework and features are obtained using Fisher vector encoding which transforms variable length low-level motion features to a mid-level representation that is more discriminant similarly to bag-of-words approaches. Contrary to countermeasures based on face/background motion consistency, the proposed method focuses only on face motion and can be applied for spoofing attacks that do not fake the whole view. We evaluate the robustness of the proposed method on the three databases (ReplayAttack-DB, CASIA-FASD and MSU-MFSD) and highlight the main limitations of non-cooperative motion-based countermeasures. Particularly, the

evaluation of motion-based countermeasures is new for the MSU database and we demonstrate that good detection is achieved despite of camera shakes when a mobile sensor is employed.

Our fourth contribution investigates the recapturing process and proposes a parametric model to describe the different mechanisms involved between a real face and its recaptured version. In this model, we consider radiometric and blur distortions specifically and propose two techniques to recover the different distortions between real faces and fake ones using enrolment samples as a reference. The estimated parameters are directly used as features for classification and we prove that both distortions can be a great help for fake face detection provided that the sensor quality is good enough and that acquisition conditions between enrolment and testing are similar. However, our real motive is to investigate the consistency of these distortions from one individual to the other and we take a first step toward the synthesis of spoofing attacks for new individuals.

Finally, our contribution to the BIOFENCE project is presented. The direct application of the certification methodology based on the Common Criteria originally developed for fingerprint technology on face biometric systems does not reflect the resistance of face biometric systems against spoofing attacks as it is. Hence, propositions for the adaptation of the methodology to face anti-spoofing are made.

0.5 Thesis outline

The main objective of this thesis is the evaluation of the resistance of 2D face recognition biometric systems against spoofing attacks. After a review of spoofing techniques and protection measures, we take advantage of the available public databases and make a complete evaluation of texture-based and motion-based countermeasures on the most recent ReplayAttack, CASIA and MSU spoofing databases. We focus on non-cooperative software methods which consider the face region only in anticipation of mask attacks. Then, we explore radiometric and blur distortions in a model-based approach and take a first step toward the synthesis of spoofing attacks. This document is organised as follows.

In chapter 1, we describe the state of the art of spoofing techniques along with the recent public databases implementing these attacks. We present in detail the selected databases for this study along with their respective evaluation protocols. Then, we prove the vulnerability of unprotected 2D face recognition systems against photo, video and mask attacks. Protection measures are essential and we provide a general overview of state of the art protection methods. A particular effort is provided to highlight useful cues to detect spoofing attacks and software countermeasures are discussed in more details under the light of these discriminant cues.

In chapter 2, a detailed review of texture-based methods is presented. In a second part, an unified framework for the evaluation of texture-based countermeasures is presented and a fair evaluation of different state of the art texture descriptors on the ReplayAttack, CASIA and MSU databases is provided. At last, improvements of the traditional LBP descriptor are proposed using color and feature selection.

In chapter 3, a detailed review of motion-based countermeasures is presented. Then, we propose a novel motion-based countermeasure using rigid and non-rigid motions and perform exhaustive evaluations on the ReplayAttack, CASIA and MSU databases.

In chapter 4, a model-based approach to the anti-spoofing problem is adopted. The recapturing

process is modelled as a combination of blurring and radiometric transformations and two methods are proposed to recover these distortions using enrolment samples. In a second attempt, attack synthesis is explored using sparse coding.

In chapter 5, a draft of the certification methodology designed for fingerprint technology is presented. This methodology is then evaluated in the context of face anti-spoofing and propositions for improvements are given.

Chapter 1

State of the art

Contents

Since the first [Chakka11] and second [Chingovska13a] IJCB competitions on face anti-spoofing, the development of new public spoofing databases and protection solutions have escalated. In this chapter, we first provide a general overview of existing spoofing attacks. Then, we highlight the main ideas used for the development of protection methods against attacks using software techniques. Finally, we discuss existing databases and evaluation standards employed in the literature.

1.1 Spoofing attack forgery

Spoofing stratagems keep evolving and present new treats for in-place anti-spoofing countermeasures. In this section, we present how spoofing attempts are implemented in real situations. The first step toward designing a fake face is to obtain the face biometry of a target user. This information takes several forms and condition the type of fake face that can be made. Fake faces are classified into three categories: photos, videos and masks. The easiest way to perform an attack is to steal a photo or video of a valid client with or without his/her consent. It can be done directly by filming discretely the target client or indirectly via social networks or any on-line media sharing services in which the client is registered. This data can then be utilized to manufacture a fake face. When the user's cooperation is possible, high-quality acquisitions are obtained to manufacture a

realistic fake face. In particular, realistic masks can only be achieved by obtaining a high quality 3D scan of the client face so the client's cooperation is required. The second step is the fake face manufacturing:

- **Photo and video attacks** Photo attacks use either printed photographs or digital photos. To manufacture print attacks, the hijacked picture is printed using a high-end consumer printer on matte photo paper in order to render a realistic fake face. The quality of the reproduction is dependent of the printer, ink and paper characteristics. In practice, standard A4 printouts are able to spoof unprotected face recognition systems. For digital photo attacks, the quality of the attack is often limited by the screen characteristics. Actual digital photo attacks use the latest smartphones and iPads as displays for practicability. Similarly, video attacks are performed using smartphones or iPads. Usually video acquisitions have a lower quality than photo captures as automatic exposure and focus are set differently. The next generation of photo and video attacks are likely to use high-contrast monitors and HDR acquisitions for more realistic rendering. Figure 1.1 illustrates different photo and video attacks.

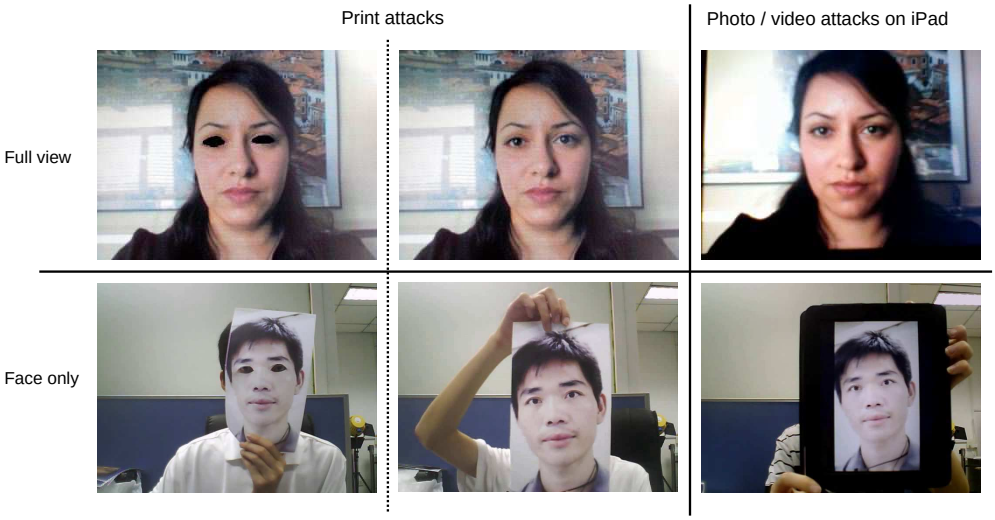


Figure 1.1: Exemplars of photo and video attacks. The first row shows examples of full view print attacks with eye-cut, basic print attacks and video attacks. The second row contains their face only counterparts.

- **Mask attacks** Various techniques exist to manufacture face masks at a wide price range. With the arrival of 3D printers on the consumer market, affordable masks can be obtained quite easily from online services like "ThatsMyFace.com" which specializes in custom face mask sculptures. Only two pictures (one frontal and one profile) of the face are needed to reconstruct a reliable 3D face model. The face sculpture takes different forms from cheap paper-craft masks (30$) to expensive real-size resin-based masks (300$) as illustrated in figure 1.2. With the target user cooperation, it is possible to manufacture a more realistic mask using a 3D scan of the face. Another technique requiring artistic skills consists in painting a silicone mould of the face.

The last step designates the way the attack is performed in front of the authentication sensor. From a given type of fake face, multiple attack scenarios are possible. In the case of mask attacks, if real size masks are employed then the impostor just needs to wear the mask and act normally in front of the sensor following the authentication procedure. When miniature masks are used, the

Figure 1.2: Exemplars of masks from ThatsMyFace.com. From left to right, pictures represent a real size paper-craft mask, a miniature mask, a real size decoration mask and a wearable mask.

impostor must hold the mask in front of the sensor and may try to simulate real face behaviour by manipulating the mask (stretching, bending, moving, ...). The displaying manner is as important as the quality of the fake face when mimicking a real face realistically. In the case of print attacks, similar manipulations are performed to simulate liveness. Another way consists in cutting some face parts such as eyes and mouth to simulate motion through the holes. For video attacks, ideally the screen is fixed so only the replayed motion is visible. In practical situations however, the screen is hand-hold in front of the sensor with more or less hand motion.

Another key aspect must be considered when referring to photo and video attacks. Full view attacks correspond to attacks that cover the whole sensor view and include a fake background in addition to the fake face region as illustrated in figure 1.1 (first row). This way, the spoofing medium is no longer visible making the attack much more difficult to detect for a real person. Face only attacks designate distant attacks that cover only the face of the impostor. In this case, the face occupies the whole support (paper or screen) and has a better resolution due to the inverse zooming effect.

The whole spoofing attack forgery pipeline is illustrated in figure 1.3. We clearly distinguish the attack type which refers to the type of fake face and the attack scenarios referring to the use of the fake face. Even though fake faces are limited to three types, progress in multimedia technologies (cameras, printers and screens) provides the impostors with new means to make more realistic fake faces. In particular, real size realistic masks become affordable with the development of 3D printers and are bound to pose new treats in a near future.

This review of spoofing attack forgery answer the first objective of the BIOFENCE project. In the case of 2D face recognition systems, spoofing attacks are already well known and the implementation of new spoofing attacks is directed toward expensive mask spoofing. Real size masks are manufactured within the BIOFENCE framework, otherwise no other breakthrough face spoofing attacks have been developed in this work. We assume that the latest public face spoofing databases are already covering a wide range of spoofing attack scenarios and we concentrate on the latest ones to evaluate the treat to 2D face recognition systems. The next section describes the evolution of public databases and their evaluation protocols.

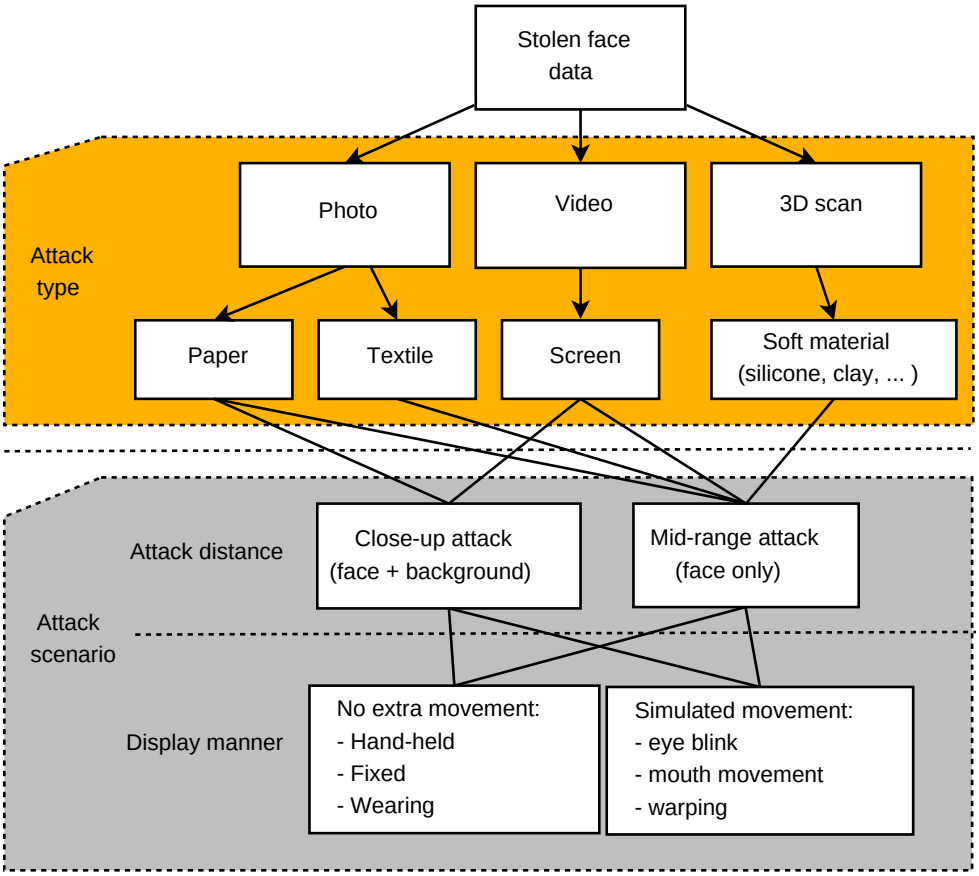


Figure 1.3: Design of face spoofing attacks.

1.2 Public databases and evaluation standards

In this section, we present the public databases used for the development of software-based face anti-spoofing countermeasures. Also, we define some mainstream evaluation schemes and connect them with current standards in the biometric community.

1.2.1 History

Fake face detection has been an active research field in the past few years and several public databases have been released for testing new anti-spoofing algorithms. In 2010, Tan and al. [Tan10a] developped the NUAA Imposter database to test 5 photo-attack scenarios with low and medium quality printed pictures under different lightning conditions. Different types of movements (translations, rotations, wrapping, bending) are recorded to test all displaying configurations in photo spoofing. Soon after, the YALE-Recaptured database complicates the detection problem with illumination variations and new attacks using digital photos displayed on LCD screens. In 2011, the Print-Attack database was released by Anjos and Marcel [Anjos11]. It provides a larger set of high resolution printed photos and real accesses along with a licit testing protocol for a fair comparison of anti-spoofing countermeasures. Photos are either hand-hold or fixed to test motion based counter-measures and videos are collected instead of pictures unlike previous databases. In 2012, the Replay-Attack database [Chingovska12] was designed as an extension to the Print-Attack database. It gathers photo and video attacks (fixed or hand-hold attacks) on different mediums (iPhone and iPad) under controlled or complex backgrounds. At the same time, a more complex

database named CASIA-FA DB [Zhang12] includes warped photo attacks to simulate facial motion, eye-cut photo attacks to simulate blink and videos attacks using an Ipad. In 2013, IDIAP researchers created the public 3D mask Attack Database [Erdogmus13]. It contains depth maps and 2D face images. Spoofing attemps were conducted with real-size masks and paper-cut masks from "ThatsMyFace.com". They also provided an update to the "Print Attack" database by adding digital photograph spoofing attacks on low and high resolution displaying mediums (see "Photo Attack" database [Anjos14]).

Table 1.1: Timeline of anti-spoofing public databases release.

2010	NUAA Imposter DB [Tan10a].
2010	YALE-Recaptured DB [Peixoto11].
2011	PrintAttack DB [Anjos11].
2012	CASIA-FA DB [Zhang12] & ReplayAttack [Chingovska12].
2013	3D-Mask DB [Erdogmus13].
2014	MSU-MFS DB [Wen15].

The mentioned public databases cover all known spoofing attacks under a user-friendly authentication protocol where the client tries to remain still and natural in front of the camera. Several self-collected databases have been created for the evaluation of interaction-based countermeasures. Especially, multiple motion-based countermeasures impose yaw and pitch head motion to assess the three dimensionality of the face during authentication. The release of these public databases has considerably boosted the development of software-based countermeasures.

1.2.2 Evaluation schemes

Recent anti-spoofing evaluation methodologies [Chingovska13b] grow toward a joint evaluation of the anti-spoofing module and face recognition/verification module, transforming the traditional two class spoofing detection problem (real accesses and attacks) into a pseudo-ternary classification problem that includes genuine access, zero-effort impostors and spoofing attempts. This type of evaluation better reflects the real performance of the biometric system in real applications as the anti-spoofing module affects the final performance of the system. In this thesis, we stick with the traditional binary classification evaluation of anti-spoofing countermeasures as it enables fair comparisons with state of the art methods.

Although some unifying efforts have been proposed in [Marcel14], multiple evaluation strategies and conventions exist in different publications. We present the main evaluation conventions associated with existing public databases.

Validation schemes Binary classification systems are evaluated using samples from real accesses and samples from spoofing attacks using a training set and a testing set of non-overlapping identities. A common practice is to further divide the testing set into a development set for model parameter tuning and a test set for reporting results. When a small amount of data is available, cross-validation evaluation strategies are employed to improve generalization.

Performance metrics Anti-spoofing methods are subject to two types of errors. Either a real access is classified as a spoofing attempt (False Rejection or False Negative) or a spoofing attack is considered as a real access (False Acceptance or False Positive). The terminology adopted in most of the publications is the following:

- False Acceptance Rate (FAR): ratio of incorrectly accepted spoofing attacks for a given threshold.
- False Rejection Rate (FRR): ratio of incorrectly rejected real accesses for a given threshold.

Performance of anti-spoofing methods is evaluated in terms of error rates (FAR and FRR). Graphical representations of error rates include Receiver Operating Curves (ROC), Detection Error Tradeoff (DET) curves and Expected Performance Curve (EPC). To report quantitative results, several metrics are proposed by selecting a given operating point. Popular metrics are:

- Equal Error Rate (EER): designates the error rate at operating point t_0 (classification threshold) where FAR equals FRR.
- Half Total Error Rate (HTER): corresponds to the mean between FAR and FRR at a given operating point. The HTER depends on the classifier threshold which is set using a cross validation set (development set). A common practice is to set this threshold so that FAR = FRR on the development set.

Industrial applications follow strict security specifications and must not exceed a maximum FAR (generally $FAR < 0.5\%$). To evaluate the performance of countermeasures according to industrial standards, the FRR at FAR = 0.1% is sometimes reported.

1.2.3 Selected databases and evaluation protocols

As discussed in the previous chapter, the number of public databases have exploded since 2010 and exhaustive experiments are now possible to assess the real potential of the proposed countermeasures. In this chapter, we describe the databases considered in this work. Experiments are conducted on ReplayAttack-DB [Chingovska12], CASIA-FAS DB [Zhang12], MSU-MFS DB [Wen15] and Morpho-MA DB [Kose13a]. These datasets have complementary characteristics. The ReplayAttack database deals with static full view close-up replay attacks whereas CASIA-FAS DB focuses on face only mid-range replay attacks with simulated motion. The MSU-MFS DB complements the ReplayAttack dataset by adding better quality print attacks and by using a mobile sensor. Finally, real-size mask attacks are tackled using the Morpho-MA DB.

1.2.3.1 IDIAP Replay-Attack Database (ReplayAttack-DB)

Using a standard web-cam, the Replay-Attack database aims to evaluate the performance of anti-spoofing countermeasures against different types of replay attacks with an increasing quality. The Replay-Attack database [Chingovska12] is publicly available at the IDIAP Research Institute website[1].

Recordings The database contains videos of both real-access and spoofing attack attempts of 50 different subjects. Attacks are performed at close-range and cover the whole sensor's view (full view attacks). The authentication process requires the users to face the sensor and to remain still for about 10 seconds. The acquisitions were carried out with a 320*240 resolution webcam of a MacBook Laptop during about 10 seconds at a frame rate of 25 fps, under two different lightning conditions. Both uniform background with artificial lighting and non-uniform background under natural illumination have been considered. Three different types of attacks were considered with an increasing level of resolution. First, *mobile* attacks are performed using photos and videos taken with the iPhone 3GS displayed in 480*320p on iPhone screen. Second, *print* attacks use plain A4 printed photos taken from a 12.1 mega pixels Canon camera. The same camera is used to record videos in 720p (HD). Both photos and videos are displayed on an iPad in 720p to perform *highdef* attacks. Those attacks are performed in two manners, hand-hold and fixed. An illustration is given in figure 1.4. Because full view attacks are performed, photo and videos used to manufacture the attacks are captured in the same conditions as real accesses.

Protocols The database is split into three sets for evaluation. A training set containing 20 subjects is provided to train the spoofing detector (binary classification). Reporting results is done by setting the detector threshold on the EER of the development set (15 subjects) and by computing the corresponding HTER on the test set (15 subjects). To investigate the efficiency of anti-spoofing countermeasures against diverse spoofing attacks, six protocols are designed:

- *print* Only print attacks are considered.
- *mobile* Photo and video attacks performed by iPhone are considered.
- *highdef* Photo and video attacks performed by iPad are considered.
- *photo* Photo attacks performed by iPhone and iPad are considered.
- *video* Video attacks performed by iPhone and iPad are considered.
- *overall* All attacks are considered.

1.2.3.2 CASIA Face Anti Spoofing Database (CASIA-FASD)

The CASIA database investigates two aspects: the impact of the sensor quality (low, medium and high resolution sensors are used) and the impact of simulated motion when detecting high quality print and video attacks. The CASIA Face Anti-Spoofing DB [Zhang12] is publicly available from the Chinese Academy of Sciences Center for Biometrics and Security Research (CASIA-CBSR)[2].

[1] http://www.idiap.ch/dataset/replayattack
[2] http://www.cbsr.ia.ac.cn/english/FaceAntiSpoofDatabases.asp

Figure 1.4: Raw exemplars of authentication attempts using client 022 identity from ReplayAttack DB. The left figure corresponds to raw samples and right figure corresponds to the corresponding face region. Top line shows acquisitions under natural illumination while bottom line display recordings under artificial lighting. From top to bottom, real accesses, print attacks, mobile attacks and Ipad attacks are displayed respectively. Only digital photo attacks are illustrated as video attacks look similar (when a single frame are displayed).

Recordings CASIA-FASD contains videos of real-accesses and replay attack attempts of 50 different subjects. The authentication process requires the users to face the sensor for about 10 seconds and movements are tolerated as long as the client looks at the camera. Photos and videos used to create fake faces are acquired with a Sony NEX-5 digital camera which produces 1080p (full HD) videos. The attacks fake only the face region and are performed at mid-range. Printed photos are made from a frame of the recorded videos and A4 copper paper for better quality. Print attacks are performed in two manners: the attacker deliberately warps an intact photo, trying to simulate facial motion (warped photo attack) or the photo is cut along the eyes and the attacker wears it like a mask while exhibiting blinking through the holes (Papercut photo attack). For video attacks, an iPad is used to display the fake faces just like the Replay-Attack database. To evaluate the impact of the sensor quality used for authentication on anti-spoofing performance, three devices with different resolutions are tested. First an old low resolution USB webcam produces low quality 640*480p video samples with faces averaging 210*190 pixels. Second, a modern USB webcam with medium resolution captures faces at about 230*200 pixels. Last the Sony NEX-5 digital camera serves as the high resolution sensor producing 710*620 face samples. Typical samples are shown in figure 1.5.

Protocols The database is split into a training set (20 subjects) and a test set (30 subjects). Results should be reported using DET curves [Martin97] and EER on the test scores. Seven scenarios are defined to evaluate the effect of imaging quality and attack types on anti-spoofing countermeasures.

- *LD*: Low quality test using samples captured by the low cost webcam

- *MD*: Medium quality test using samples captured by the modern webcam
- *HD*: High quality test using samples captured by the Sony Nex-5 camera
- *Warped*: Acquisitions of warped photo attacks are considered (from all three sensors).
- *Cut*: Acquisitions of cut-photo attacks are considered (from all three sensors).
- *Video*: Video attacks are considered (from all three sensors).
- *Overall*: All attacks are considered.

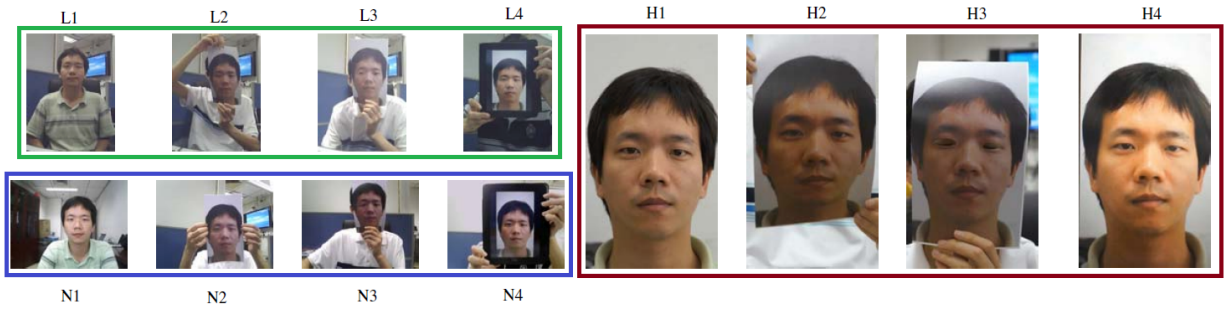


Figure 1.5: Raw exemplars of authentication attempts using client 5 identity from CASIA database. Samples are captured with low (green box), medium (blue box) and high definition (red box) sensors respectively. In each box, samples represent real access, warped photo attack, cut-photo attack and Ipad attack respectively.

1.2.3.3 MSU Mobile Face Spoofing Database (MSU-MFSD)

The MSU database [Wen15] investigates the problem of face spoofing in mobile phone applications (mobile phone unlock) using high quality full view replay attacks. This database was produced at the Michigan State University Pattern Recognition and Image Processing (PRIP) Lab and is publicly available on demand[3].

Recordings MSU-MFS DB contains videos of real-accesses and replay attacks of 35 different subjects. The authentication process requires the users to face the sensor and to remain still for at least 9 seconds. Attacks are performed at close-range and cover the whole sensor's view similarly to those in ReplayAttack DB. Two types of sensors are used for authentication, the built-in camera of a MacBook Air 13" and the front facing camera of the Google Nexus 5 smartphone. Three types of attacks are conducted using different supports. Photographs are printed on A3 paper using a HP Color Laserjet CP6015xh printer from 18Mp photographs recorded using a Canon PowerShot 550D SLR camera. The same camera also captures 1080p video clips that are displayed on an iPad Air screen (2048*1536 pixels) to generate iPad video attacks. Another type of video attack is conducted using an iPhone 5S, with screen resolution 1136*640, for capturing and displaying 1080p fake face videos. Attacks are performed using a fixed support. Examples are shown in figure 1.6. Because full view attacks are performed, photo and videos used to manufacture the attacks are captured in the same conditions as real accesses.

[3]http://www.cse.msu.edu/rgroups/biometrics/Publications/Databases/MSUMobileFaceSpoofing/index.html

Protocols The database is divided into a training set of 15 subjects and a test set of 20 subjects. Seven scenarios are defined to evaluate the effect of image quality and attack types on anti-spoofing countermeasures.

- *Android*: Samples captured by the Google Nexus 5 smartphone.
- *Laptop*: Samples captured by the built-in webcam of a MacBook Air 13".
- *Print*: Acquisitions of print attacks are considered (from both sensors).
- *iPhone*: Acquisitions of video attacks displayed on iPhone are considered (from both sensors).
- *iPad*: Acquisitions of video attacks displayed on iPad are considered (from both sensors).
- *Video*: Acquisitions of cut-photo attacks are considered (from both sensors).
- *Overall*: All attacks are considered.

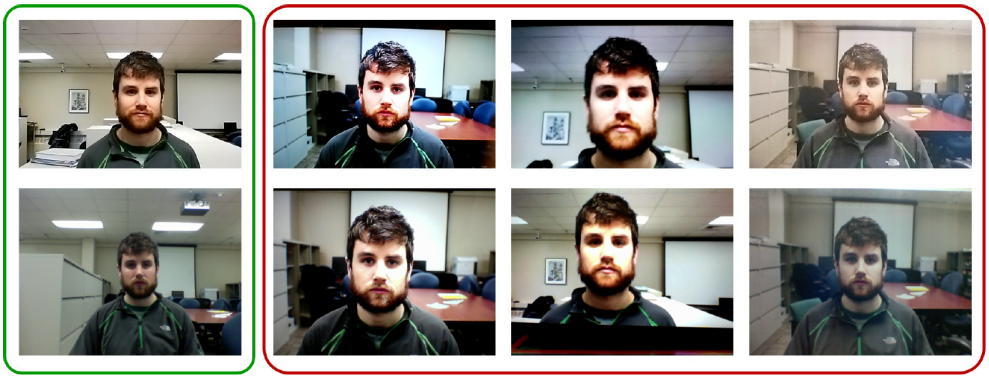


Figure 1.6: Raw exemplars of authentication attempts of one of the clients in the MSU-MFS database. Samples are captured using Google Nexus 5 smartphone (first row) and a MacBook Air 13" laptop camera (second row). Columns corresponds to: (a) real accesses, (b) iPad video attacks, (c) iPhone video attacks and (d) printed photo attacks.

1.2.3.4 Morpho Mask Attack Database (MorphoMAD)

MORPHO-MAD investigates the problem of mask attack detection using high quality real-size masks more realistic from a texture standpoint than those from the 3DMask Attack database. This database is a proprietary database accessible only to partners of MORPHO[4].

These masks are colorless and were designed for spoofing 3D face recognition systems. It contains 2D IR acquisitions (gray-scale photos) of both real-access and mask attack attempts of 20 different subjects for whom 16 masks have been manufactured. The remaining 4 subjects without mask complete the database to have more real-access samples. For a given mask, 12 attack attempts are carried in total by making several subjects wear the mask under different poses including the owner of the mask. To obtain realistic masks, a 3D scanner which uses a structured light technology captures face shape characteristics of the target person. Then the 3D mesh is derived from the projection of the acquisition into a polygon 3D model and sent to the 3D printer. Masks were manufactured by Sculpteo 3D Printing. Some examples are given in figure 1.7.

[4]http://www.morpho.com/en

Figure 1.7: Raw exemplars of authentication attempts of one subject in the MORPHO-MAD database. (a) Genuine user, (b) Attack performed by the owner of the mask, (c) Attack performed by an impostor.

1.2.4 Evaluation schemes

The public databases have their own evaluation protocols. For Replay-Attack database, three data subsets are provided for training, tuning and testing. The classification threshold is obtained from the development set at the EER then, using this threshold, results are reported in terms of HTER on the test set. For CASIA and MSU databases, only two sets are provided for training and testing and results are reported using DET or ROC curves and EER.

1.2.5 Summary

The recent multiplication of public databases has boosted the development of a large panel of countermeasures. The present work falls within this context and new software-based methods are developed on three of the most challenging spoofing databases publicly available: ReplayAttack-DB, CASIA-FASD, MSU-MFSD. We highlighted the complementarity between each dataset. Table 1.2 summarizes relevant properties of each database. Unfortunately, the Morpho-MAD contains photo acquisitions in near infra-red as the masks are in black and white. Consequently, it is inconsistent with RGB video recordings of the considered public databases for which we have proposed motion and color-aware countermeasures. For this reason, mask attacks are only considered to demonstrate their versatility against 2D face recognition systems and experiments on RGB mask attack recordings are left to future work.

Table 1.2: General properties of the selected databases.

Databases	Replay-Attack DB [Chingovska12]	CASIA-FASD [Zhang12]	MSU-MFSD [Wen15]	MorphoMAD [Kose13a]
# subjects	50	50	35	20
# samples	1200	600	280	392
Authentication protocol	still	slight movements (expression changes, pose variations)	still	little pose variations but eye-closed
Illumination	adverse controlled	adverse	adverse	controlled
Sensor	Built-in webcam MacBook 13" (320*240)	• Low cost webcam (640*480) • Standard webcam (480*640) • Sony NEX-5 (1920*1080)	• Built-in camera in MacBook Air 13" (640*480) • Nexus 5 Android phone (720*480)	NIR camera (480*640)
Impostor camera	Canon PowerShot SX150: - 12.1 Mp photos - 720p video at 30 fps	Sony NEX-5: - 1080p video at 25 fps	• Canon PowerShot 550D SLR: - 18 Mp photos - 1080p videos • iPhone 5S: - 1080p videos	3D scanner
Attack type	• print photo • photo on iPhone • photo on iPad • video on iPhone • video on iPad	• warped photo • cut photo • video on iPad	• print photo • video on iPad	silicone mask
Attack distance	close-up, full view	mid-range, face only	close-up, full view	mid-range
Display manner	hand-held & fixed	hand-held with simulated motion	fixed	wearing the mask with closed eyes

1.3 Vulnerability of 2D face recognition systems against fake faces

In order to evaluate the vulnerability of unprotected 2D face recognition systems against spoofing attacks, we investigate multiple face verification use-cases under the treat of state of the art spoofing attacks. Intrinsically, face recognition systems are resistant to spoofing attacks to some degree. For instance, if the fake face is too different from the enrolled real access sample due to geometric distortions or low quality recapture the authentication attempt is viewed as belonging to a different identity and is rejected. Our goal is to assess how easy it is to fool a face verification system using state of the art spoofing attacks. First, the face recognition algorithm used in our experiments is described. Then, the evaluation of its resistance against various spoofing attacks is investigated under several use-cases.

1.3.1 Face recognition algorithm

The architecture of the face recognition algorithm is presented in figure 1.8. First, faces extraction is performed using the Pittpat 5.0.2 SDK. Eyes are located and a face registration procedure geometrically aligns the face so that eyes are horizontal. Extracted faces are cropped and resized to 128*128 pixels. Then, face images are converted into gray-scale before performing illumination corrections using Tan and Triggs preprocessing scheme [Tan10b]. A baseline face recognition algorithm based on Gabor features and Principal component analysis is used to represent face images into a suitable space for matching. The matching is performed using a nearest neighbours classifier and the cosine Mahalanobis distance. In our experiments, the system is used in verification mode meaning that a one-to-one matching procedure compares the input sample with the claimed identity template registered in the system during the enrolment phase. The implementation is based on the Matlab toolbox[5] provided by Vitomir Struc [Štruc10].

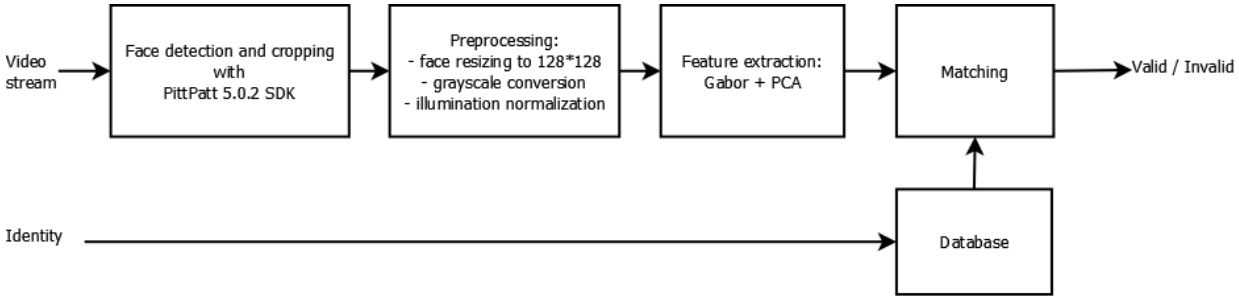


Figure 1.8: Face recognition pipeline.

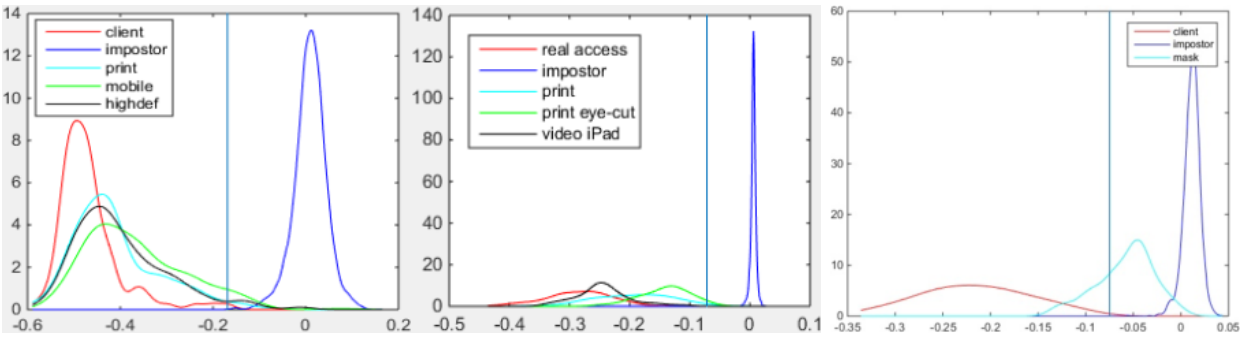
1.3.2 Evaluation of the resistance of 2D face recognition towards spoofing attacks

Our goal is to assess the permeability of unprotected systems against fake faces. Attacks from the ReplayAttack, CASIA and Morpho databases are considered to investigate most of existing attack scenarios. First, experiments on close-up replay attacks are conducted using the ReplayAttack database. Second, mid-range attacks from the CASIA database are considered. Finally, mask attacks from the MorphoMAD are evaluated. Because different acquisition conditions are used to capture the face from one database to the other, the face recognition algorithm is tuned on the training set (gallery) for each database. Only one image per identity is used to build the gallery in our experiments and also a single query image per identity is considered during the deployment phase (verification phase). For each identity claim, three different cases occur:

- the real client corresponding to the claimed identity checks in (real access).
- an impostor checks in (zero-effort attack).
- an impostor checks in with a fake face corresponding to the claimed identity (spoofing attack).

Hence, we analyse the matching score for each of these scenarios. Only one real access attempt and one spoofing attack attempt are tested for one identity claim whereas N-1 zero-effort attacks are tested where N denotes the number of clients enrolled in the face recognition system database.

[5] https://fr.mathworks.com/matlabcentral/fileexchange/35106-the-phd-face-recognition-toolbox



(a) ReplayAttack-DB (b) CASIA-FASD (c) MORPHO-MAD

Figure 1.9: Distribution of face recognition scores. The vertical line corresponds to the matcher threshold.

The decision threshold to accept or reject an authentication attempt is set so that there is the same error rate between false rejections and false acceptances with respect to real access and zero-effort attacks.

Use-case1: close-up attacks from ReplayAttack-DB The ReplayAttack database contains video recordings of real accesses and close-up replay attacks which cover the whole view hiding the spoofing medium borders. Only one frame is extracted from the video to perform the face verification. This database contains a specific set for building the gallery of identity templates during the enrolment phase. The distribution of the matching scores for real access, impostor attempts (also referred as zero-effort attacks) and spoofing attacks are reported in figure 1.9a. The face recognition system obtains almost perfect recognition performance as the $EER = 0.01\%$. However, the spoofing attack success rate is very high with 98% for print attacks, 94% for mobile attacks and 96% for iPad attacks.

Use-case2: mid-range attacks from CASIA-FASD The CASIA database contains video recordings of real accesses and mid-range replay attacks which hide the face of the impostor. Only one frame is extracted from the video to perform face verification. As this database does not contain an enrolment set, another frame is extracted to build the gallery. The distribution of the matching scores for real access, impostor attempts (also referred as zero-effort attacks) and spoofing attacks are reported in figure 1.9b. Similarly to the ReplayAttack case-study, the face recognition system obtains perfect recognition performance as the $EER = 0.01\%$ and the spoofing attack success rate is very high with 98% for print attacks, 98% for print eye-cut attacks and 100% for video iPad attacks.

Use-case3: Mask attacks from Morpho-MAD This database contains pictures of authentication attempts of real access and realistic mask attacks using a near infra-red camera. The distribution of the matching scores for real access, impostor attempts (also referred as zero-effort attacks) and mask attacks are reported in figure 1.9c. The system recognizes perfectly each individual but fails to detect efficiently mask attacks as 26% of them bypass the system.

1.3.3 Discussion

Photo, video and mask attacks present a real treat to unprotected face recognition systems. Especially, photo and video attacks are very easy to implement both in terms of skills and resources. Mid-range attacks and close-up ones obtain more than 94% chances of success and demonstrate the urgent need of protection measures against spoofing attacks. Mask attacks are more difficult to manufacture and the success rate is not as good as photo and video attacks but they are expected to raise challenging problems from an anti-spoofing perspective in the near future as they become more and more realistic with the rapid development of 3D printing technology.

1.4 State-of-the art in face anti-spoofing

Face recognition has been an active field for a long time but face anti-spoofing research has only captured the attention of the biometric community until the last decade following the success of fingerprint biometrics. Nonetheless, extensive work has addressed the problem of securing face authentication systems and main research directions have been established.

1.4.1 General taxonomy of anti-spoofing strategies

The general classification of anti-spoofing strategies is depicted in figure 1.10. Countermeasures are divided into hardware-based and software-based methods.

- **Software-based** methods process the data collected from the authentication sensor. Dynamic approaches detect voluntary or involuntary motions of the face region to discriminate real and fake faces while static approaches focus on texture analysis and image quality assessment.
- **Hardware-based** methods use extra hardware for the anti-spoofing task in addition to the sensor used for recognition. These methods are further divided into liveness measurements, attack specific detection and challenge-response approaches. Liveness measurements category regroups techniques that employ a specific sensor to detect particular attributes of living bodies such as body temperature, blood flow, electric pulse or skin reflectance. It also includes multi-biometric approaches which rely on the verification of the identity from multiple biometrics (face, voice, fingerprints, iris, ...). Attack specific detection refers to methods designed against a particular type of spoofing attack. For example, photo and video attacks can be detected from depth measurements obtained with a depth camera, stereo vision or multiple focus measurements. Challenge-response approaches designate methods that rely on the cooperation of the user for answering a random request during authentication. For example, the system may request the user to move his/her head following a motion pattern generated randomly during the authentication. The fact that the request is unpredictable makes it hard to spoof.

General comparisons of different classes of methods are drawn from main criteria:

- Performance: the system must match industrial specifications in terms of security. Depending on the application, a maximum false acceptance rate is imposed (usually below 0.5%) to make

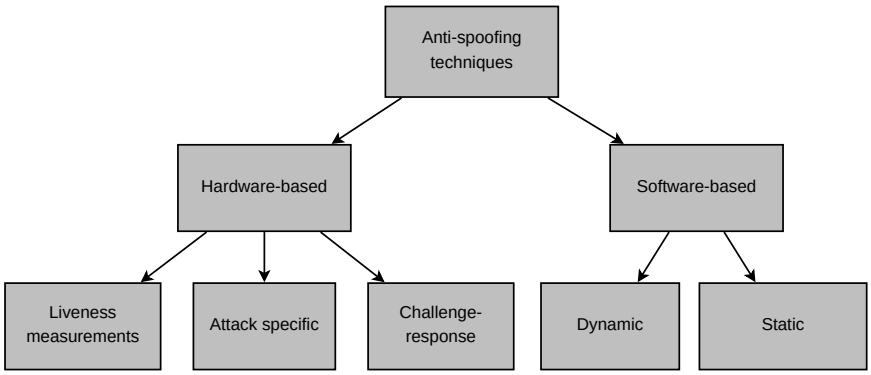


Figure 1.10: General classification of anti-spoofing strategies.

sure that almost all the attacks are detected. The anti-spoofing module must satisfy this constraint while maintaining good recognition performance for practicability.

- Cost: deployment of the anti-spoofing solution and the recognition system must stay in the range of reasonable cost for the considered application.
- User friendly: interaction with the system should be easy and fast enough.
- Genericness: the system must be applicable to many use-cases. It is robust to acquisition conditions such as scene context and illumination changes but also can cope with unseen attacks.

While hardware-based methods offer the best security guarantees in general, extra costs are generated by adding new equipment for the anti-spoofing task. Besides, hardware-based solutions are usually not user friendly and generic as it requires specific settings to work properly. For example, challenge-based methods involve user-system interactions during a significant amount of time. Also, methods relying on multi-spectral and infra-red acquisitions need specific illumination conditions and are sensitive to make-up. On the other hand, software-based approaches usually achieve lower performance than hardware-based methods but remain cheaper as no extra equipment is needed. Their integration into existing recognition systems is easier and more generic as they apply to standard RGB sensors. In this work, we focus on software-based methods to take advantage of existing spoofing databases presented in section 1.2.3.

1.4.2 Discriminant cues

To understand the motivations behind the exhaustive set of software-based countermeasures, we first identify the underlying cues behind each method. Going through the literature, eight types of cues have been exploited so far. The diagram in figure 1.11 depicts the nomenclature of cues considered for software-based anti-spoofing. Static cues are displayed in green and dynamic cues are shown in red.

- **Quality loss** refers to perceived image degradation due to the recapturing process of a spoofing attack. It takes several forms and we identify three main incarnations. First, exposure and color shifts can be observed as different sensors are employed for fake face manufacturing and authentication. Second, as printers and screens have limited performance, limited resolution of fake faces induces a noticeable lack of details when the authentication sensor is

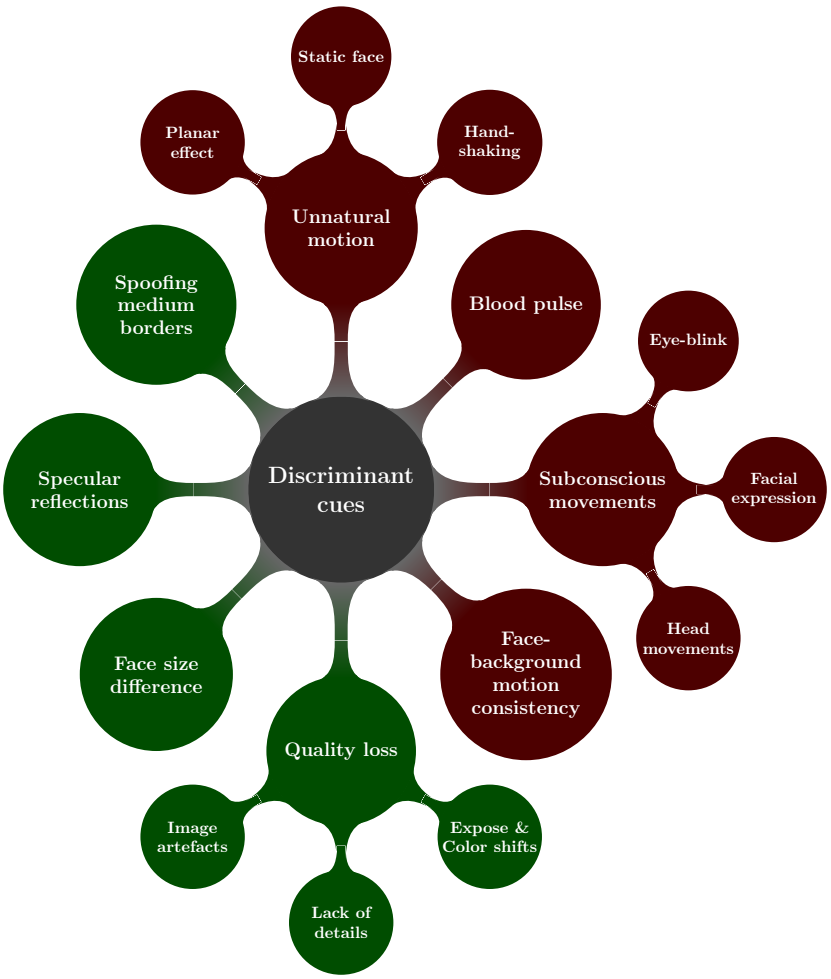


Figure 1.11: Taxonomy of discriminant cues. Static cues are displayed in green and dynamic cues are shown in red.

of good enough quality. Last, image artefacts are present due to the different procedures of capturing, compressing, manufacturing (printing or displaying) and recapturing. This type of cues is very specific to a given authentication set-up (sensor and acquisition conditions dependent) but also to particular attack scenarios.

- **Face size difference** happens when performing full view attacks for photo and video attacks. As papers or screens have limited size, usually fake faces appear bigger from the sensor point of view as the spoofing medium gets closer to hide its borders. This can be circumvented easily by selecting the right view when manufacturing the attack as it is possible to anticipate the distance to the sensor required to display the attack and maintain the same face size between real accesses and spoofing attempts.

- **Specular reflections** are different between real faces and fake ones for several reasons as the displaying support (paper, screen, mask) used for spoofing inherits different optical properties than skin. Besides, photos and screens can have additional ambient reflections due to their planar glossy surface. This type of cue is present regardless of the attack type but it highly depends on illumination conditions.

- **Spoofing medium borders** are visible when mid-range face only photo or video attacks are performed.

- **Unnatural motion** designates uncanny movements of the face region when performing photo or video attacks. When holding the fake face or the camera, some characteristic hand-shaking motion may be captured. Conversely, attacks can be displayed on a fixed support and the absence of motion indicates that photo attacks are probably attempted. Another type of unnatural motion stems from the planar geometry of photo attacks that limits out-of-plane head movements such as yaw and pitch head motions. Unfortunately, this type of discriminant cue depends on the attack scenario and can be partly countered by scenarios where the user manages to mimic real head movements successfully through training.

- **Subconscious movements** correspond to the counterpart of the previous cue as the focus is on natural motion instead of unnatural aspects. Living faces express involuntary movements routinely such as eye-blinking, facial expression changes and subtle head movements due to the respiratory rhythm. These movements are subtle yet highly discriminant against photo attacks. The advantage of this type of cues is that it is independent to acquisition settings (sensor type, illumination conditions, authentication protocol) but it is weak against video and mask attacks.

- **Blood pulse** can be measured by single video acquisitions of the face using photoplethysmography. This type of cue is very generic and discriminate real faces from all type of spoofing attacks. However, fixed illumination conditions and a long authentication time ($> 8s$) is required for a good estimation as multiple heart beats are required. Recent advances in this field has led to real-time commercial solutions[6] for heart rate estimation but its use for real-time liveness detection is yet to be demonstrated.

- **Face-background motion consistency** is observed on full view photo attacks as the face region and the background are part of the same physical image. It is robust to varying acquisition settings but it is very specific to this type of attack scenario and can no longer be used against mid-range or mask attacks.

1.4.3 Taxonomy of software-based methods

Software-based methods are divided into static and dynamic approaches in [Galbally14b]. In [Tirunagari15], the authors suggest another taxonomy of software-based countermeasures to connect with the underlying cues behind each countermeasure. Software-based methods are divided into model-based (cue-based) and data-driven methods. The first category relies on the aforementioned set of cues to detect real and fake faces. These cues are determined either by observation or intuition when analysing fake/real faces and they are not always consistent between two different spoofing scenarios. Besides, the precise identification of robust discriminant cues is difficult and most of the works in the literature resort to data-driven methods for texture and motion analysis. This second class of methods rely on low level differences between real and fake faces that are extracted using generic descriptors and supervised classification. The separation between the two classes is not well defined and can be seen as the line between high-level methods relying on strong priors about the differences between real and fake faces (cue-based) compared to low-level methods motivated by blind assumptions on the existence of differences at texture level or motion level. This categorization facilitates the comprehension of the motivations and assumptions behind existing countermeasures to better assess their strengths, their limitations and their complementarity.

A general categorization of software-based countermeasures following this taxonomy is presented

[6] `http://www.i-virtual.fr/`

in figure 1.12. Pioneer works proposed countermeasures relying on either motion aspects (in green) or texture aspects (in red). Since the first IJCB competition [Chakka11] on face anti-spoofing, system experts have been introduced by combining motion and texture cues. In parallel, another type of approaches (in yellow) jointly considers dynamic and static cues and proposes spatio-temporal descriptors for extracting discriminant features.

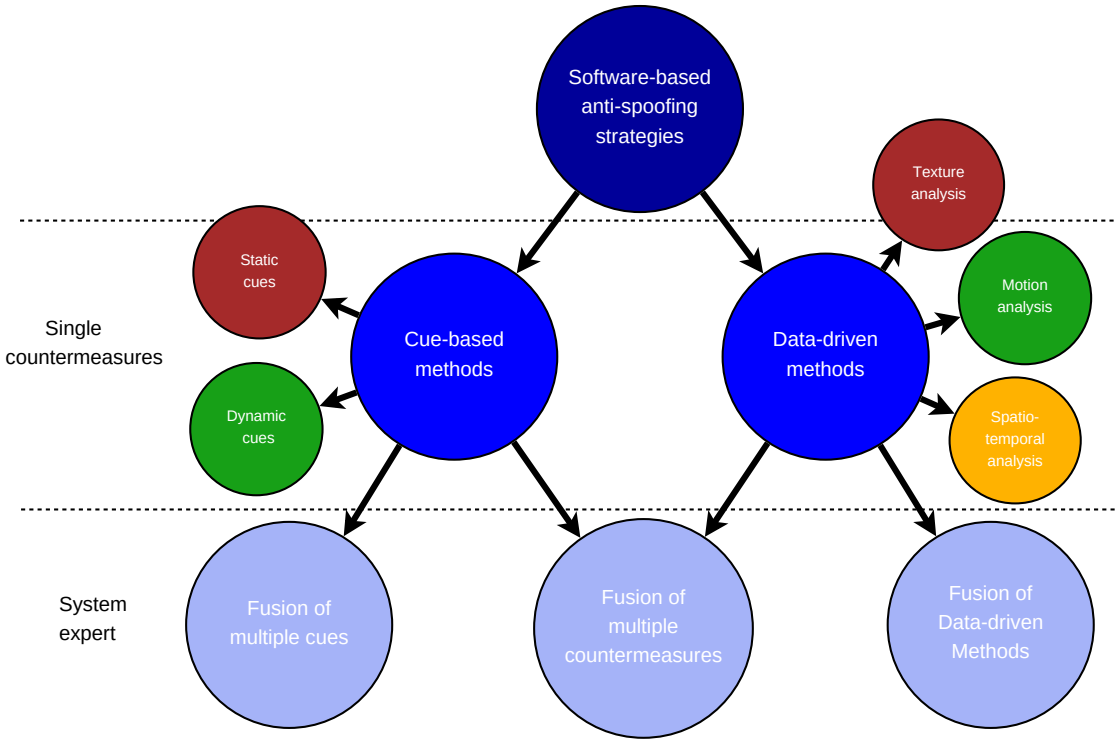


Figure 1.12: High level classification of software-based anti-spoofing strategies.

The implementation of software-based countermeasures is usually independent of the face recognition algorithm. It is often placed before the face recognition stage because in the case of a detected fake face the recognition stage is irrelevant. In this set-up, both anti-spoofing and recognition performance are assessed independently and may not reflect the true performance of the joint system. Indeed, the joint task of anti-spoofing and recognition (verification) is a ternary classification problem where real clients, impostors (zeros-effort attack) and spoofing attacks must be properly labelled. In [Chingovska13b], the authors investigate the fusion of the recognition and anti-spoofing modules at decision-level and score-level to handle correctly the discordant responses of both modules (the recognition phase rejects impostors but the anti-spoofing module accept them). In [Chingovska15], the same authors propose to first perform the recognition and then use client-specific anti-spoofing detection based on the claimed identity to handle spoofing attacks. Perfect recognition performance is assumed as fake face detection is relevant only in this case anyway. In this thesis, we suppose that this assumption holds as recent face recognition systems outperform humans for the verification task with the development of deep learning architectures. Consequently, we can reasonably assume that spoofing attacks represent the only source of failure of face biometric systems and only the performance of the anti-spoofing module is relevant. We consider the anti-spoofing task only after identification and the anti-spoofing module is placed after the recognition module.

1.5 Conclusion

In this chapter, we described the general methodology to forge fake faces and invent different spoofing scenarios. We proved that a large variety of spoofing attacks are already implemented in existing public spoofing databases and we demonstrated their high versatility towards unprotected 2D face recognition systems.

A brief overview of anti-spoofing strategies has been presented along with some insights on discriminant cues between real accesses and spoofing attacks. The way the attack is performed greatly influence the realism of the attack from a motion and image quality standpoint, especially close-up and mid-range attacks raises different challenges when designing countermeasures as discriminative cues are specific to each attack type and scenario. This work focuses on software based methods and takes advantages of the recent release of ReplayAttack, CASIA and MSU spoofing databases publicly available. These databases are selected for the evaluation of new countermeasures developed in the course of this work.

In order to cope with the large variety of attack scenarios and in view of more advanced mask attacks, only the face region can be considered for fake face detection. Both static and dynamic cues must be exploited. Radiometric and blur distortions associated with quality loss and specular reflections discriminant cues are the most consistent among various spoofing attacks and are key to the development of new countermeasures proposed in this thesis. Dynamic cues are incapable of dealing with video or mask attacks except blood pulse estimation. Nonetheless, the use of dynamic cues can help to increase the robustness of texture-based countermeasures to deal with certain type of attack scenarios. For this reason, we investigate the potential of motion-based countermeasures on a complete set of attack scenarios to draw out their strengths and limitations.

Chapter 2

Countermeasures based on texture analysis

Contents

Following the two IJCB competitions in 2011 and 2013 [Maatta11] [Chingovska13a] many countermeasures have been proposed. Recent methods have converged toward the fusion of complementary cues involving motion and texture aspects. The contribution presented in this chapter is threefold. First, a complete state of the art of static texture-based methods is provided to get a deep

understanding of existing countermeasures for the development of new ones. Second, this study aims to unify existing results on texture-based anti-spoofing methods across the recent databases under a common evaluation framework that deals with face region only and which takes advantage of high resolution acquisitions. Third, given the success of LBP-based methods in anti-spoofing, texture classification and face recognition tasks we investigate two LBP variants. Motivated by our analysis of the recapturing process presented in chapter 4, we propose to use contrast and color information to improve the classic LBP features. We demonstrate that the color texture countermeasure based on HSI-LBP features outperforms current data-driven texture-based methods when only the face region is considered and fair well against recent cue-based methods based on image quality assessment (IQA) [Galbally14a] or image distortions analysis (IDA) [Wen15].

2.1 State of the art of texture-based countermeasures

Methods based on texture analysis are key methods in video anti-spoofing as they do not require any user-cooperation during the authentication process nor any additional equipment to detect photo and videos attacks. Furthermore they are usually low computational and they can be computed on a frame by frame basis for a fast response. With the arrival of video and mask attacks, these methods became essential in the development of recent countermeasures. Following the proposed nomenclature of anti-spoofing strategies in Chapter 1, we present a complete review of texture-based countermeasures in two parts: cue-based and data-driven methods. The frontier between cue-based methods and data-driven methods is not absolute and some countermeasures can be viewed from both aspects. We chose to group generic texture analysis methods relying on generic texture descriptors under the data-driven class of methods. In that regard, methods such as Local Binary Patterns (LBP) are discussed as belonging to the data-driven category although they are also used to characterize noise cues (banding effects, printing artefacts, ...). First, we describe the main discriminant cues coming out from the literature and from our expertise. Then, cue-based and data-driven countermeasures are presented. The proposed review is quite incomplete as dynamic approaches are not discussed in this work, apart from a few exceptions, although most recent countermeasures (in 2014 and 2015) tend to exploit spatio-temporal information.

2.1.1 Discriminant texture-cues

As mentioned in Chapter 1, discriminant static cues that are addressed in the literature are: quality loss, face size difference, specular reflections and visible support borders. We believe that face size differences between real and fake faces and the visibility of the spoofing medium are cues that are not consistent enough with the recent advances in spoofing attack forgery. Full view attacks hiding the medium borders outside the sensor view are easily performed and with proper expertise one can make sure that the recaptured face has the same size as its real version by adjusting the viewing distance and size of the fake face. For these reasons, we only focus on specular reflections and quality loss cues in this work.

2.1.1.1 Specular reflections

Spoofing attacks use a 2-D planar support either photo paper or digital screens. As a consequence, some uncanny specular reflections are sometimes observed on fake faces (see figure 2.1). Besides,

captured digital attacks manifest some saturation effects at specular highlights due to some over-exposure as screens are direct light sources.

2.1.1.2 Quality loss

Quality loss is a perceptual term designating all sorts of image degradations. We categorize these distortions into three categories.

Exposure and color-shifts The recapturing process generates some exposure and color changes between a real face and its recaptured version under the same illumination conditions. A complete study of the recapturing process is presented in Chapter 4.

Lack of details and blur Blurriness is observed on close-up spoofing attack scenarios for two main reasons. First, the limited size of screens act as a low pass filter when resizing the high resolution face data for display. Besides, the smaller the screen the closer it is from the sensor to perform full view attacks (covering the whole scene to mask the support boundaries) and acquisitions tend to be out of focus.

Noise Also, the generation of spoofing attacks involves several procedures of capturing, compressing, displaying and recapturing. Capturing refers to the process of conversion of photons arriving at the camera sensor into electrical charges. Those charges are then transformed into digital information to form the final image. This procedure generates two types of noise in images that can be identified and used for camera identification in [Luka06]: the fixed pattern noise (FPN) and the noise resulting from the photo-responsiveness of non-uniform light-sensitive cells (PRNU). So recaptured scenes have peculiar noise compared to natural ones. Furthermore, digital contents are usually stored into a lossy format such as JPEG or Bitmap that introduces additional noise. Furthermore, in case of print-attacks, printing generates image artefacts such as banding, jitter and ghosting as described in [Eid11]. For digital attacks, monitor screens also produce undesirable effects such as distortion, flickering and Moiré patterns during the recapturing process.

Figure 2.1 illustrates the aforementioned distortions with exemplars extracted from public databases.

2.1.2 Cue-based methods

Only 3 works tackle the anti-spoofing problem from a cue-based approach. First, the work of Wen and al. [Wen15] proposed to detect three types of image distortion characterized by abnormal specular highlights, additional blurriness and color distortions. They propose four different features to account for these unnatural effects. Specular reflection features are derived from the specular reflection removal method of [Tan05] which was already used in the general image recapture detection method of [Gao10]. Even though the extraction of the specular component has a hard time separating the diffuse part and the specular part on face images, informative features can still be derived for the detection task. Blur features are derived from two blind (no-reference) perceptual blur metrics that were proposed in [Crete07] and [Marziliano02]. Finally, the last two color features

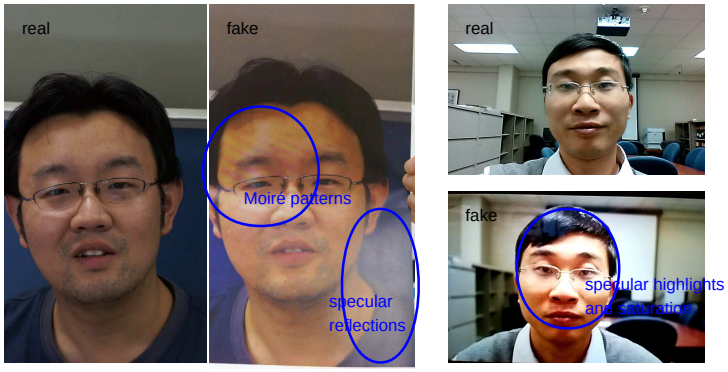


Figure 2.1: Illustration of different distortions between a real face and a fake one. Moiré patterns and specular reflections are highlighted on an exemplar from CASIA database. Abnormal specular highlights with saturated values are illustrated on an exemplar from the MSU database. Note the overall exposure color distortions between real and fake faces. Blur is difficult to perceive due to different colors.

characterize the color distribution of the face in terms of histogram moments (in the HSV color space) and color diversity.

Second, Pinto and al. [Pinto12] used visual rhythm and GLCM on the Fourier spectrum of the noise residual video to capture noise generated by fake face. Particularly, Moiré patterns are considered so the whole image is taken into account as major artefacts occur in flat regions outside the face. Visual rhythm consists in reorganising the 3D video content into a 2D image to have a compact representation for which classic image processing can detect noise in both space and time. To extract the noise information, a Gaussian filter is employed on each frame of the video to produce a noise free video from which a simple frame by frame difference with the original video provides a noise residual video containing discriminative noise information. Then, the noise pattern is extracted from the low-band in the Fourier spectrum of the video for each direction (horizontal and vertical) as notable differences between real and fake face are observed. Thus, central horizontal lines and central vertical lines are extracted to represent each frame in the visual rhythm procedure. GLCM computed on this 2D noise signature signal is performed and features are fed to an SVM for classification. This method works perfectly on their private database containing LCD screen photo and video spoofing attempts on controlled environment because strong Moiré patterns are present in recaptured images. Lower results are obtained on the Replay-attack database as they achieved $HTER = 15.62\%$ because less distortions are perceptible for this database.

Third, the method proposed by Galbally and al. in [Galbally14c] investigates general image quality assessment (IQA) metrics chosen for the face anti-spoofing problem. Quality properties have already been explored for liveness detection in fingerprints and iris applications. Human can feel the difference between recaptured face and genuine ones. The goal of IQA in this case is to quantify with objective metrics a reliable estimation of "appearance" perceived by humans. A lot of metrics exist in order to measure the degree of sharpness, color and luminance levels, entropy, structural distortions, contrast, level of digital compression, ect. IQA techniques designed for anti-spoofing must combine several of those to capture effectively the appearance differences between real and fake faces. Metrics have been selected from broadly used methods that revealed good performance for different applications and sustain complementary properties of the image such as contrast, texture (entropy), sharpness. Only low complexity metrics have been used. A total of 25 quality metrics have been investigated by team ATVS in the face anti-spoofing TABULA RASA challenge on the Replay-attack database [Galbally14c],[Chingovska13b].

2.1.3 Data-driven methods

Texture-based data-driven methods employ texture analysis approaches to characterize the differences between real and fake faces and learn a discriminative model in a supervised manner. We distinguish two broad classes of texture-based countermeasures. The first one uses spectral texture analysis methods and the second one statistical approaches.

2.1.3.1 Spectral approaches in texture analysis

Spectral approaches refer to the frequency domain where features are related to statistics of filter responses. Evidence shows that a tuned bandpass filter bank resembles the structure of the neural receptive field in the human visual system. Feature extraction in spatial frequency domain has several advantages. First, a filter is selective as it enhances only certain features while suppresses others. Second, the periodic structure of a texture can be explicitly represented in the spectral domain.

Spectral methods usually use filter banks or image pyramids to convert an image from the spatial domain into the frequency domain and vice-versa. Like in statistical methods, the distribution of feature measures (e.g., wavelength coefficients for a wavelet transform) provides a sparse texture description which can be used directly as input for further classification. However, the resulting description is over-complete, because it contains an increase and thus redundancy in information content.

Methods based on the Fourier transform In [Li04], Li et al. addressed the print attack issue by analysing high frequency of face images and its frequency dynamics with 2D Fourier transform. Small size fake images have less high frequency components compared to real face images. The authors built the high frequency descriptor (HFD) defined as the ratio of the energy of high frequency bands over the total energy of the face image. Also, frequency variability is lower in case of photo attacks because expressions and poses of the face remain invariant even with a moving picture. Therefore another discriminative descriptor can be derived by computing the standard deviation of the HFD for a short video sequence. This method works well on their self-collected database which contains down-sampled printed photo attacks but is likely to fail for high quality pictures.

Reflectance analysis with band-pass filters In [Tan10a], Tan et al. discard the specular reflection and assume that face images are made solely of diffuse light. By using the Lambertian model, Tan is able to extract the reflectance and the illumination components by using either a Logarithm Total Variation method (LTV) or a Difference of Gaussian (DoG) approach. Their intuition is that reflectance of 2D fake faces must be uniform compared to 3D real faces as it depends exclusively on the surface orientation. Then, classification is achieved with a Sparse Low Rank Bilinear Logistic Regression directly on the reflectance component derived from DoG or LTV. Their method works well even on gray images acquired from a standard webcam. In [Peixoto11], Peixoto and al. (Unicamp) extended Tan and al. algorithm to deal with images under challenging illumination conditions on the same database and improved the classification accuracy by 6.6%. They suggest a pre-processing step using CLAHE (Contrast Limited Adaptative Histogram Equalization) to reduce illumination changes.

The variational retinex framework is also used in [Kose13b] by Kose and Dugelay to detect mask attacks on Morpho database. Unlike Tan and al, Kose used the variational retinex algorithm [Gross03] [Kimmel03] process to derive the reflectance component. They proved that reflectance is a better discriminative feature than the most popular LBP texture descriptor on this database.

In [Zhang12], Zhang et al. introduced the CASIA database together with a baseline method using Dog filtering along with an SVM classifier.

Reflectance-based methods extract the discriminative information from the high middle frequency band. Several methods have been investigated such as LTV (logarithmic total variation), DoG (difference of gaussian), AS (anisotropic smoothing) and MSR (multiscale retinex). Both DoG and AS give promising results while being simple and fast to compute. The main drawback is the lack of robustness to illumination variations, CLAHE preprocessing is able to correct this flaw partially.

Exploiting the specular component structure Inspired by the work of Yu and al. [Yu08], Bai et al. detect high definition photos attacks (digital and printed ones) by analysing the specular component of an image in [Bai10]. The authors use the bidirectional reflectance distribution function (BRDF) and the dichromatic reflection model to extract the specular component and the diffuse one from color images with Tan's algorithm [Tan05]. They prove that displaying mediums used in photo attacks have different specular responses depending on their surface texture. For print-attacks, paper granular structure and ink deposition patterns are visible in the specular component. For LCD photo-attacks, pixel grid and image encoding differ from genuine face images. Instead of comparing directly specular values, the authors propose to normalize the specular component with intensity values to enforce invariance to exposure changes. The histogram of the gradient of the normalized specular component follows a Rayleigh distribution whose shape is characteristic of a recaptured texture and a real capture. The distribution parameters are used as features and fed to an SVM classifier. Great performance is achieved on their self collected database but their method only applies to high-resolution acquisitions that can resolve fine texture patterns. Also, specular decomposition is sensitive to illumination conditions.

2.1.3.2 Statistical approaches in texture analysis

Statistical approaches collect image signal statistics from the spatial domain as feature descriptors. Usually lower-order image statistics, particularly first- and second-order statistics, are exploited in texture analysis. First-order statistics, such as the mean, standard deviation and higher-order moments of the histogram, relate to the distribution of pixels whereas second-order statistics also account for the spatial inter-dependency or co-occurrence of two pixels at specific relative positions. Grey level co-occurrence matrices (GLCM), grey level differences, autocorrelation function, and local binary pattern (LBP) operator are the most popular second-order statistics for texture description. Higher than second-order statistical features have also been investigated, but the computational complexity increases exponentially with the order of statistics and are less popular for this reason.

LBP-based methods The generalized definition of LBP from [Ojala02] is used. To encode neighbourhood relationships around a given pixel p_c, N samples points p_i with $i \in \{1, N\}$ distributed

evenly on a radius R around p_c are used to compute the associated lbp code as:

$$lbp_{N,R}(p_c) = \sum_{i=1}^{N} s(p_i - p_c)2^i \text{ , where } s(p) = \begin{cases} 1, & \text{if } p > 0 \\ 0, & \text{otherwise} \end{cases} \quad (2.1)$$

The normalized histogram of the lbp codes forms the classic LBP features:

$$LBP_{N,R} = hist(lbp_{N,R})/N$$

where N is the number of pixels in the LBP image.

LBP-based methods are the most popular and the most successful texture-based countermeasures for anti-spoofing. In face anti-spoofing, many authors chose to use only the uniform patterns as it provides a good trade-off between efficiency and cost. Non uniform patterns are grouped into a single bin of the histogram. Uniform patterns are LBP codes with at most 2 bitwise transitions 1-0 or 0-1 in their bit representation. It was shown in [Ojala02] that most of the patterns in natural images are uniform and considering only uniform patterns adds statistical robustness while reducing significantly the number of possible patterns from 2^N to $N(N-1)+3$. It is also possible to compute a spatial version of LBP refereed as $LBP_{N,R,b*b}$ features by dividing the image into $b*b$ blocks and concatenating LBP features from each block into a single feature vector.

Multiple works use the LBP descriptor combined with an SVM classifier with a radial basis function (RBF) kernel for fake face detection. The first authors to popularize this approach for fake face detection are Maata and al. in [Maatta11]. They argue that better results are achieved by extending the classic LBP approach to a multi-scale representation (MS-LBP) and using a 3*3 multi-block LBP scheme leading to a 833 feature vector. On the NUAA database, they demonstrate the superiority of LBP over spectral methods such as Gabor features or Tan and al. approach [Tan10a] and this method becomes a baseline for later public databases. The LBP descriptor has been successfully employed to detect print attacks from the ReplayAttack corpus and achieved perfect detection on the first IJCB face anti-spoofing challenge [Chakka11]. The face region and its surroundings are considered for better success as full view print attacks are evaluated (the whole image is fake).

Later, Kose and al. implemented MS-LBP as a baseline method for 3D mask antispoofing on their self-collected mask attack database (Morpho-MAD) in [Kose13a, Kose13c]. Also, Chingovska and al. evaluated the performance of MS-LBP on the face region only for the recent Replay-Attack database and CASIA database in [Chingovska12]. Their study draw out two main observations. First, non linear SVM classification outperforms the traditional ξ^2 Nearest Neighbour classifier and the Linear Discriminant Analysis classifier (LDA) on both ReplayAttack and CASIA databases. Second, multi-block LBP computation obtains similar results compared to traditional LBP on the ReplayAttack database and using the high dimensional MS-LBP approach yields only little improvements for this database and for CASIA database. Also, a set of extended LBP was investigated: transitional (tLBP), direction-coded (dLBP) and modified LBP (mLBP) as described in [Trefny10] but no significant improvement was obtained. Finally, the traditional LBP countermeasure serves as a baseline on the 3D Mask attack database by Erdogmus [Erdogmus13].

In [Kose12], Kose and Dugelay took advantage of contrast and texture information using rotation invariant local binary patterns variance (LBPV) to detect fake faces. They prove the superiority of LBPV compared to LBP on the NUAA database. They also proposed DoG filtering as a viable pre-processing step to reduce misleading information such as noise due to illumination variations. This preprocessing step is similar to the retina-based filter presented by Ngoc Son Vu in [Son Vu10].

The Local Binary Patterns from the Three orthogonal Planes (LBP-TOP) is the natural extension of LBP to spatio-temporal texture analysis. By taking into account temporal variations, significant improvement is achieved compared to static LBP features. We mention this descriptor as it defines the new baseline for anti-spoofing evaluations. In [Komulainen13], the authors evaluate the LBP-TOP features when only the face region is considered and obtain almost perfect results on the CASIA database and perfect results on the print attacks from the ReplayAttack database. Complementary experiments are conducted on the ReplayAttack database in [Freitas Pereira13] where the face region is rescaled to 64*64 pixels. Performance is slightly worse than those of the previous study on the CASIA database.

Low level image descriptors In [Tronci11], Tronci and al. used low level visual features usually employed for content based image retrieval to detect print attacks from the ReplayAttack database and have obtained perfect detection during the first IJCB challenge. Numerous features are employed to have a dense description at different visual levels covering texture to color aspects. A multi-classifier system provides a score for each type of feature and a dynamic score combination methodology is used to combine all these scores into a final decision. Their method takes advantage of the whole image and additional motion and liveness countermeasures.

In [Schwartz11], Schwartz and al. also use low level descriptors including color frequency, histogram of oriented gradients (HOG), GLCM features and histograms of shearlet coefficients. An extended face region is considered and faces are resized to 110*140 pixels for their experiments on the PrintAttack database. All these features are fed to a partial least square discriminant analysis (PLS-DA) for classification. The proposed low level features perform well individually except for HOG and their combination yields slight performance improvements with an EER decrease of around $3-4\%$.

2.1.4 Fusion of statistical and spectral methods

In [Waris13], the authors evaluated three well known texture features, namely $LBP^{riu2}_{16,2}$, Gabor features and GLCM Haralick's features on the ReplayAttack database. The whole face is considered and per-frame features are averaged to form a single vector for each video. Gabor features are obtained using the mean and the standard deviation of the magnitude of the transform coefficients (4 scales and 6 orientations). Classification is performed using an SVM classifier and a Partial Least Square discriminant analysis method (PLS-DA). Better results are actually obtained with the PLS-DA and $LBP^{riu2}_{16,2}$ and Gabor features outperform GLCM features. They proposed to concatenate both features and achieve perfect detection on the ReplayAttack database.

Kim proposed to exploit frequency and texture information to detect print-attacks in [Kim12]. He proposed a 1D power spectral density feature vector by sampling the 2D power spectrum ($|FFT|^2$) into 32 concentric rings. Texture features are derived from the popular $LBP_{8,1}$ descriptor. Both feature vectors are then concatenated and fed to an SVM classifier.

In [Yang13], Yang and al. proposed a component-based method to resolve the scaling issue inherent in texture analysis. Using the local parts in the H-face representation enables a better focus on discriminative texture patterns. The framework consists in 4 steps: locating the components of face, coding the low-level features (Local Phase Quantization, Local Binary Patterns, Histogram of Oriented Gradients) respectively for all the components, deriving the high-level face representation by pooling the codes with weights derived from Fisher criterion and concatenating the histograms

from all components into a classifier for identification. Top results were achieved on NUAA, Print-attack and CASIA database.

2.1.5 Overview

From this dense state of the art on texture-based countermeasures, we draw out some relevant observations that guide our work. On the one hand, cue-based methods are easier to interpret and generalizes better on other databases compared to data-driven methods. However, finding consistent discriminant cues with respect to attack scenarios and sensors is not an easy task and most cues are very sensitive to illumination variations. On the other hand, data-driven methods handle multiple static cues blindly directly by learning a discriminative model from generic texture descriptors but has the disadvantage to be data specific (bad generalization to other database). Tables 2.1 and 2.2 recapitulate texture-based countermeasures.

Another point of emphasis is the success of LBP features across public databases. MS-LBP and LBP-TOP variants have proved to be powerful tools to extract discriminant texture information between real and fake faces. Along this line, non linear SVM is also widely used for classification although linear classifiers such as PLS and LDA demonstrated similar performance when high dimensional features are employed.

Besides, it appears that the region of interest for exploiting texture information is crucial. While perfect detection is achieved when the whole scene is considered, detection performance drops significantly when only the face region is used. Various face extraction and preprocessing schemes have been explored but, to the best of our knowledge, no study has really discussed the impact of these procedures on the detection performance.

Additionally, recent works combine multiple texture descriptors but also complementary motion or liveness cues. While significant improvements are achieved using this type of strategy, it becomes harder to evaluate the contribution of its canonical constituents for further improvements. The general methodology in this thesis is to isolate texture and motion-based countermeasures and to build on novel strategies to improve on both aspects separately. This strategic bias contrasts with the recent line of research where development of spatio-temporal countermeasures is investigated. However we believe that it allows a better understanding of discriminant information for fake face detection.

Table 2.1: Summary of texture-based anti-spoofing countermeasures. The column 'Attacks' corresponds to the spoofing attack scenarios handled by the proposed countermeasures. Column 'ROI' indicates which part in the image is considered for the countermeasure. Column 'Database' mentions the name of the database used to validate the countermeasures. For specific works, we mention in parenthesis the name of the team during the first or second IJCB competition that are associated.

Cue-based				
Reference	Methodology	ROI	Attacks	Databases
2012, Pinto and al. [Pinto12, Pinto15b, Pinto15a] (Unicamp)	Exploit noise residuals (Moiré patterns) in the spectral domain	face + background	photo and video	ReplayAttack DB, Proprietary DB
2014, Galbally and al. [Galbally14a] (ATVS)	Use general image quality assessment metrics	face + background	photo and video	ReplayAttack DB, CASIA DB
2015, Wen and al. [Wen15]	Exploit specular distortions, blur and color distortions	face only	photo and video	ReplayAttack DB, CASIA DB, MSU DB

Table 2.2: Summary of texture-based anti-spoofing countermeasures. The column 'Attacks' corresponds to the spoofing attack scenarios handled by the proposed countermeasures. Column 'ROI' indicates which part in the image is considered for the countermeasure. Column 'Database' mentions the name of the database used to validate the countermeasures. We mention in parenthesis the name of the team under which the authors have participated in the first or second IJCB competition.

Data-driven methods				
Reference	Methodology	ROI	Attacks	Databases
2004, Li and al. [Li04]	Exploit high frequency energy and its variation over time using the FFT	face only	print	Proprietary
2010, Tan and al. [Tan10a]	Exploit the reflectance component extracted with LTV and use SLRBLR for classification	face only	print	NUAA DB
2010, Peixoto and al. [Peixoto11]	Use Tan and al. method with CLAHE as preprocessing to deal with bad illumination conditions	face only	print	NUAA DB
2013, Kose and al. [Kose13b]	Exploit the reflectance component extracted with the variational retinex algorithm + linear SVM	face only	masks	MorphoMAD
2010, Bai and al. [Bai10]	Analyse the gradient distribution of the specular component	face only	print	Proprietary
2011, Schwartz and al. [Schwartz11]	A dense set of low-level descriptors is fed to a PLS-DA classifier	extended face region	print	PrintAttack DB, NUAA DB
2011, Tronci and al. [Tronci11] (AMILAB and PRALAB)	Low-level texture-based countermeasures are combined using a dynamic score combination scheme	face + background	photo and video	ReplayAttack
2011, Maatta and al. [Maatta11]	Study MS-LBP, LPQ and Gabor features with SVM-RBF classification	face only	print	NUAA DB
2012, Chingovska and al. [Chingovska12]	Study LBP variants with SVM-RBF, LDA and χ^2 classification	face only	photo and video	ReplayAttack
2013, Kose and al. [Kose13a, Kose13c]	Use of MS-LBP with linear SVM	face only	masks	MorphoMAD
2012, Kose and al. [Kose12]	Use of LBPV on preprocessed image via DoG with global matching for rotation invariance and χ^2 classification	face only	print	NUAA DB
2013, Freitas and al. [Freitas Pereira13]	Use of LBP-TOP and SVM-RBF	face only	photo and video	ReplayAttack
2014, Komulainen and al. [Komulainen13]	Use of LBP-TOP and SVM-RBF	extended face	photo and video	PrintAttack DB, CASIA DB
2013, Yang and al. [Yang13]	Exploit the face structure and use codewords derived from LBP, LPQ and HOG features	extended face region	photo and video	NUAA, CASIA, PrintAttack DB
2014, Waris and al. [Waris13] (MUVIS)	Use Gabor, GLCM and LBP features with SVM or PLS-DA classifier	face + background	photo and video	ReplayAttack DB

2.2 Definition of an unified framework for texture-based countermeasures design

Multiple strategies have been adopted for the design of texture-based countermeasures. First, several regions of interest (ROI) have been considered for extracting texture features. Sometimes,

the whole scene is considered for the detection of full view attacks from the ReplayAttack database in [Tronci11, Pinto12, Waris14, Galbally14a] but in most cases only the face region is used. In [Yang13], the authors use a holistic face description based on the face contour region together with the eyes, nose and mouth regions. Besides, as far as face region is concerned various rescaling values are considered throughout the literature. Commonly, faces are geometrically normalized to 64*64 pixels but a few works chose to either keep the original face size [Komulainen13] or use a higher scale [Yang13].

Second, static texture descriptors are computed in a frame per frame basis but more stable results can be obtained by using multiple frames when video samples are available. In [Chingovska12], all video frames are considered as independent samples for classification whereas 50 per-frame features are accumulated to form one feature vector per video in [Komulainen13]. In [Wen15], the authors use a majority vote procedure to obtain a final decision from per-frame classification scores. Only one frame per video is used in [Yang13].

Third, classification schemes are divided into general classification and multi-attack classification. General classification use one classifier to detect all types of attacks whereas in multi-attack classification, an ensemble of classifiers is employed to detect each type of attacks and the resulting classification score is obtained by combining the response of each classifier as in [Wen15].

Due to this large panel of countermeasure designs, the comparison of texture descriptors is difficult. In this thesis, we propose to evaluate texture descriptors under an unified evaluation framework to assess which properties are relevant for fake face detection.

Background information represents a great help for fake face detection as it can be used either to detect directly the spoofing medium when visible or to discriminate the real background from its copy when the scene is known. However, depending on the context of use, the background may change randomly making texture unpredictable. It is also easy to imagine people doing attacks where the background is no longer informative like with mid-range attacks, cut-photos, masks or make-up. Hence, to deal with all kinds of attack scenarios in this comparison framework, we consider that the only information available for fake face detection is the face area. In this section, we investigate the different evaluation strategies using the LBP descriptor and propose a consistent framework to evaluate texture-based countermeasures exploiting the face region only. The general processing pipeline of texture-based methods focusing on the face region is illustrated in figure 2.2.

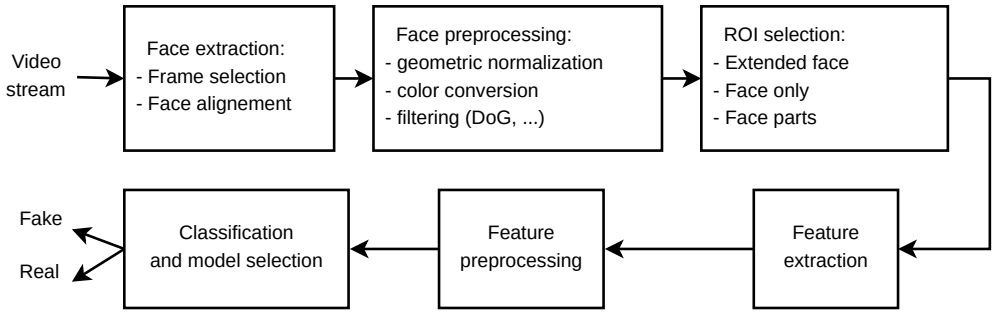


Figure 2.2: General architecture of texture-based countermeasures focusing on the face region only.

The following points are investigated to determine the best evaluation framework from a texture standpoint:

- Which face region should be considered for texture analysis?
- How to geometrically normalize extracted faces to bring out discriminant texture information?

- Is one frame enough to exploit texture information considering that texture is a static cue?
- Which feature normalization and classification settings are best?

To answer these questions, we use an advanced face tracking technique from the Openface toolkit based on facial landmarks to extract the face region. The main benefits of using this advanced face tracker are:

- Face regions are extracted at the same scale defined by the interocular distance.
- Faces are aligned to a frontal face template which corrects slight pose variations during authentication. This reduces the texture variability from one frame to another and encourages texture analysis on a single frame.
- A useful face representation is available enabling an easy segmentation of face parts.

2.2.1 Presentation of the face extraction procedure

The main problem to be tackled here is: can we derive a robust texture-based countermeasure using a single frame? The main advantage of such a design is the fast response of the countermeasure for real time authentication. Also, it reduces significantly the computation overload for our experimentations to answer the aforementioned questions. To this end, we take advantage of recent advances on facial landmark detection to extract the face region using the Openface toolkit [1]. This procedure has been used to derive motion features in chapter 3 and we propose to use it for our texture-based countermeasure design.

2.2.1.1 Face registration using the Openface toolkit

The main goal of face registration is to limit the texture variability due to pose variations (to a certain degree) and to make sure that the extracted face region is identical from one identity to the other. The rigid transformations between a standard frontal face and the observed face are computed using the Constrained Local Neural Fields (CLNF) framework for facial landmark tracking (for more details please refer to the original paper [Baltru16]). Hence, face alignment is easily performed by inverting the rotations and translations estimated by the CLNF algorithm. The scaling factor is kept to one to maintain the original face resolution unchanged. Face extraction is then performed by cropping an extended face region as defined in figure 2.3. The bounding box (in green) containing the facial landmarks is extended to make sure that face contours are encapsulated inside the extended face region. In [Yang13], the authors demonstrate that discriminant texture is localized near face boundaries. As most attack scenarios use a rectangular fake face, the immediate background region is still part of the fake face and contains discriminant texture cues. The advantage of detecting facial landmarks is that it is possible to segment only the face region efficiently by computing the convex hull of the set of facial landmarks. Exemplars of the face registration procedure with only the face region are shown in figure 2.4. Another advantage of this face extraction procedure is that face resolution can be measured in a robust manner from the inter pupillary distance (IPD) easily obtained from eye landmarks positions.

[1] https://github.com/TadasBaltrusaitis/OpenFace/wiki

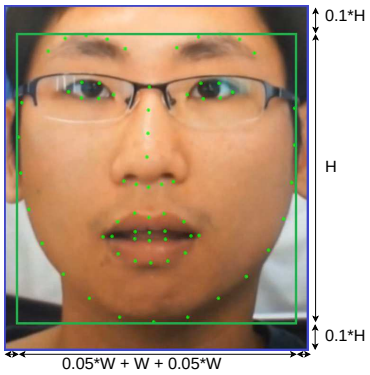


Figure 2.3: Face region extraction procedure

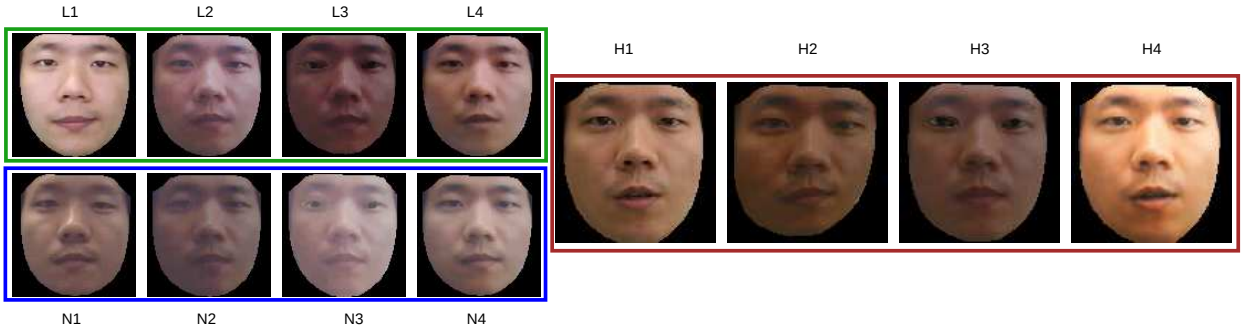


Figure 2.4: Extracted faces from the CASIA database. Aligned faces are resized for display purposes.

2.2.1.2 Frame selection

Static texture analysis is performed on a per-frame basis. Variations in the face texture between frames are very limited if the face stays with the same pose so only one frame per authentication acquisition is used to derive texture features in this work. The selected frame is picked to meet the following requirements in a hierarchical order:

- Face landmarks are correctly detected
- Eyes are opened
- Face is in frontal position
- Face motion is limited
- Face resolution is maximum

Using a single frame allows a fast response for the fake face detection and can be implemented for authentication systems working with still images.

2.2.1.3 Limitations

The face detector is not error free. Table 2.3 shows the number of wrong detections for ReplayAttack, CASIA and MSU databases. Some miss detections occur when the face covers the whole scene

and gets too close to the image boundaries. This problem can be managed partially by padding the video with border values. Another error factor is the face geometric deformation when performing attacks which lead to wrongly detected landmarks as shown in figure 2.5. Those video samples are considered as irrelevant attacks because they don't even bypass the face extraction module and they are discarded.

Table 2.3: Face detection errors

Nb of errors	real	fake
ReplayAttack	0	56/1000
CASIA	0	1/450
MSU	0	0

Figure 2.5: Exemplars of face detection errors from the ReplayAttack database

2.2.2 Face geometric normalization

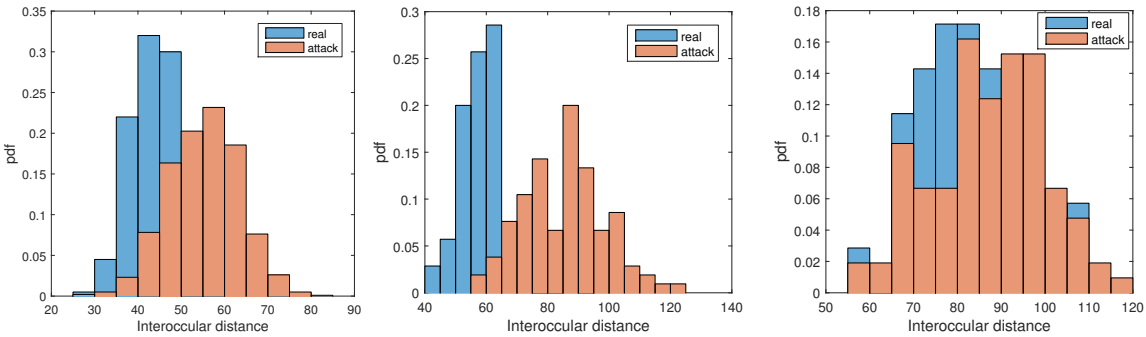
Here we want to determine the best image resolution for fake face detection using texture. The point is to define the right spatial level for texture analysis using the LBP case study.

2.2.2.1 Motivations

Face normalization is used to standardize the scale under which texture features are computed to get rid of face size variability between two authentications. Although face size differences can be discriminant between real and fake faces especially in the case of full view close-up attacks which tend to have a larger size, this cue is highly inconsistent between two attack scenarios and is discarded in this study. To highlight this observation, the distributions of face sizes for both real and fake faces for the ReplayAttack, CASIA and MSU databases are reported in figure 2.6 and 2.7. Full view attacks from the ReplayAttack database tend to have greater sizes than real faces to cover the sensor whole view properly. The same observation can be made for laptop acquisitions from the MSU database. However, the impostor can manufacture a fake face with the correct zoom so that the face region occupies the same space as a real face on the sensor view as in CASIA recordings. Besides, in practice, variable acquisition distance makes this cue non discriminant as intra-class variability increases. For example, real and fake face android acquisitions from the MSU database have similar distributions.

2.2.2.2 Experiments: LBP scaling versus image scaling

The most common practice is to normalize cropped faces (square bounding box) to 64*64 pixels before computing LBP features. This normalization is inherited from pioneer works on the NUAA



(a) ReplayAttack (b) MSU laptop (c) MSU android

Figure 2.6: Face size distribution of ReplayAttack and MSU databases.

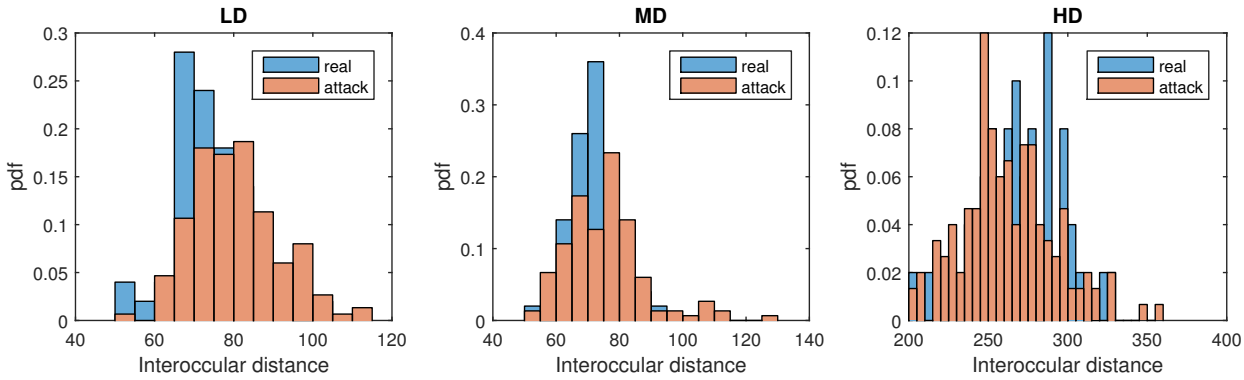


Figure 2.7: Face size distribution of CASIA database. From left to right, low quality acquisitions (LD), medium quality acquisitions (MD) and high quality acquisitions (HD).

database and existing baselines are evaluated under this evaluation framework. In [Komulainen13], the authors avoid geometric normalization to keep the original texture quality when high resolution sensors are employed. In [Wen15], faces are aligned and normalized to 144*120 pixels with an inter pupillary distance (IPD) of 60 pixels. Existing works on texture-based methods did not address explicitly the impact of geometric normalization when computing texture features. In particular, we believe that resizing the face to 64*64 pixels leads to a significant loss of information and dismisses some of the benefits of using high quality acquisitions. For this reason, we evaluate the impact of the geometric normalization with respect to the quality of the authentication sensor through experiments using the LBP features.

We investigate two strategies for the geometric normalization procedure. The first one simply normalizes face images to the average resolution of faces to globally preserve the original image quality of faces. The optimal neighbourhood configuration (radius and number of sampling points) for the LBP computation is determined by grid search. The second strategy directly resizes face images to an optimal size determined by exhaustive search beforehand and computes LBP features using the classic 3*3 neighbourhood configuration. Both strategies search for the optimal scale for describing texture to capture discriminant cues between real and fake faces.

Optimal radius strategy Experiments are conducted on the CASIA database using LBP features and an SVM classifier with a Gaussian kernel and default parameters (C =1). The three subsets corresponding to low, medium and high quality recordings are considered to evaluate how the image resolution impacts the detection. Faces are geometrically normalized so that the inter

pupillary distance (IPD) equals 80, 75 and 262 pixels for low, medium and high quality acquisitions respectively which corresponds to the average face size of each data subset. We search for the best LBP radius parameter r from the following set of values: $R \in [1, .., 5]$. When a larger radius is employed, the number of sampling points N should be increased for a detailed description of the texture but we limit the maximum number of sampling points to 16 due to the high computation overload attached. Results are reported in table 2.4. The best performance for all three subsets is achieved using $N = 16$ and $R = 5$ suggesting that coarse scales are better than fine scales for extracting discriminant LBP codes especially for high resolution acquisitions.

Table 2.4: EER results (in %) for fake face detection based on LBP features for different radius and sampling parameters on CASIA database.

N	R	CASIA LD (EER)	CASIA MD (EER)	CASIA HD (EER)
8	1	21	17	13
16	2	23	20	17
16	3	23	20	13
16	4	26	17	7
16	5	21	17	7

Rescaling strategy Faces are geometrically normalized so that the IPD ranges from 30 pixels to 270 pixels then classic LBP codes are computed within a 3x3 neighbourhood. Detection results are reported in function of the IPD values in figure 2.8. EER varies significantly in function of the IPD reinforcing the idea that correct geometric normalization is essential for the evaluation of texture-based countermeasures. The EER for the low quality sensor increases when up-sampling the face images with cubic interpolation ($IPD > 80$ pixels) showing that texture becomes less discriminant at finer scales due to the limited resolution. Surprisingly, this effect does not appear for medium quality acquisitions ($IPD > 75$ pixels) and the EER remains constant close to the minimum value. Reversely, when down-sampling the face image, better performance is achieved close to 54 IPD value for all three sensors. This goes along with the previous observation where coarser texture is more discriminant than fine texture details. Nevertheless, having a high resolution sensor helps with the detection as the minimum EER gets lower when a better sensor is used. Even though the full resolution of the high quality acquisitions is not exploited when deriving LBP features, the appropriate scale for texture countermeasures is fixed around IPD = 54 pixels. We also experiment on the MSU and ReplayAttack database to complete our analysis of the proposed rescaling strategy and similar observations are drawn. Good performance is achieved when the IPD is around 54 pixels. The optimal face rescaling value is fixed to IPD = 54 pixels when computing LBP codes within a 3*3 neighbourhood.

Table 2.5 recapitulates the average face dimensions of different subsets of the ReplayAttack, CASIA and MSU databases.

2.2.2.3 Discussion

Both experiments have led to the same conclusion: coarser scales are more discriminant than finer scales when analysing the facial texture for anti-spoofing purposes regardless of the resolution of the sensor. We believe that fine texture details are highly identity specific and attack specific which increases the intra-class variance in the binary detection problem. We now compare the

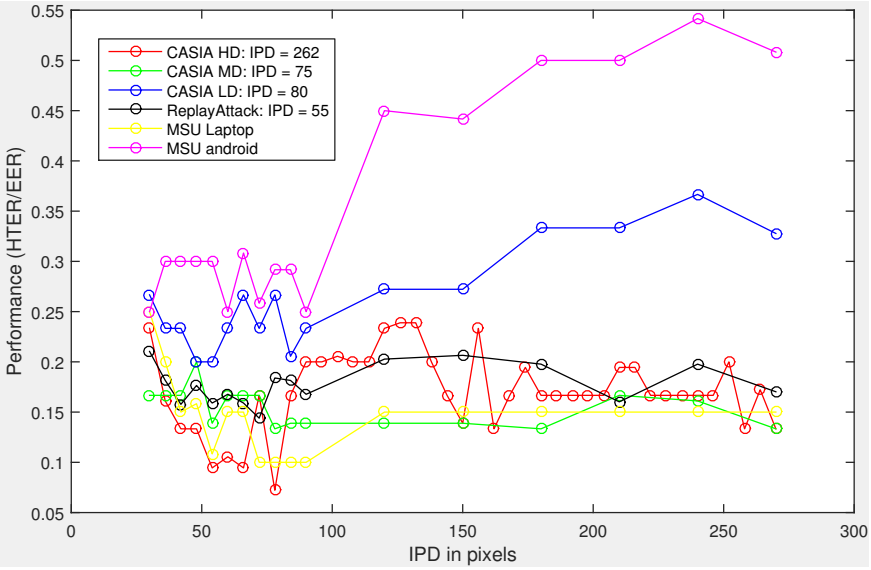


Figure 2.8: Performance of LBP features in function of the face scaling for the CASIA, ReplayAttack and MSU datasets. The IPD mentioned in the legend corresponds to the average IPD for the corresponding dataset.

Table 2.5: Average face size for the considered datasets.

Datasets	IPD	Width	Height
CASIA LD	80	190	210
CASIA MD	75	180	200
CASIA HD	262	640	700
ReplayAttack	55	120	140
MSU laptop	80	190	210
MSU android	85	210	230

results from both strategies in table 2.6. Both strategies obtain comparable results on low quality recordings. The rescaling strategy performs slightly better on medium quality acquisitions and reversely the radius tuning strategy achieves better results on high quality recordings. We retain the rescaling strategy for the rest of our experiments as the computation cost decreases thanks to the down-sampling of face images. Furthermore, this strategy set the face size to a fix empirical value regardless of the sensor used for authentication and no parameter tuning is required.

Table 2.6: Comparison of the rescaling strategy and the optimal radius strategy.

EER (in %)	CASIA LD	CASIA MD	CASIA HD
Rescaling strategy + LBP(8,1,u2)	20	14	10
Radius tuning strategy + LBP(16,5,u2)	21	17	7

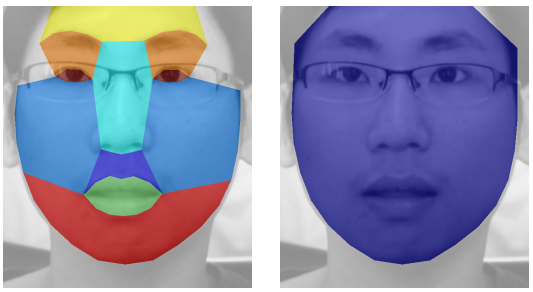
In conclusion, face texture must be computed at the same scale for all faces to ensure that the differences between two texture features reflect the quality deterioration of the recapturing process instead of the face size variability. Geometric normalization is performed on registered faces to compare properly the same facial structures from one face to the other. This procedure improves the discriminative power of LBP features in two manners. First, the standardization of

face sizes improves the consistency of features between two attempts and reduces the intra-class variance. Second, discriminant texture information is well captured at IPD = 54 pixels regardless of the sensor quality even when it leads to important down-sampling. If too much down-sampling occurs, fake face detection performance starts to degrade. In that regard, existing evaluations of texture based methods using a normalized 64*64 face image are not optimal and new evaluations are necessary.

2.2.3 Component-based face representation

The main question to be answered is: are some facial regions more relevant than others in order to detect fake faces? Should we consider the face as a whole or as a series of discriminative regions?

In the literature, multiple facial regions have been considered for texture feature extraction. To have a dense texture representation which handles various discriminant abilities of scene regions, two strategies have been investigated. The first one "spatialize" low-level features using a block processing approach while the second one takes into account the canonical structure of faces with a component-based coding. In [Yang13], the authors demonstrate that the face contour region holds important discriminant texture information compared to inside regions. Unfortunately, even near background information should be discarded to have a background independent countermeasure. The proposed face detection algorithm allows a semantic segmentation of the face and enables us to investigate texture in different coherent face regions as shown in figure 2.9a. Experiments are conducted on the ReplayAttack, CASIA and MSU databases to evaluate the discriminative power of face parts. In addition, we compare the case where the whole face is considered and the case where features from each face part are concatenated forming a component-based LBP feature.



(a) Face parts (b) Face region

Figure 2.9: Face extraction

Detection results are reported in table 2.7. The discriminative power of certain face parts depends on the image quality. For instance, cheeks are highly discriminant for low and medium quality acquisitions but not so much for high quality acquisitions. Reversely, the forehead and eye regions are discriminant for high quality samples but not for lower quality samples. Overall, the best performance is obtained using the whole face directly. Experiments using component-based LBP features on high quality acquisitions obtain $EER = 20\%$ which is 4% less than when using the whole face. The high dimensionality of the component-based LBP features probably requires more training data for better generalization. In conclusion, the whole face should be considered when deriving texture features.

Table 2.7: EER results (in %) for different face parts on CASIA database

Face parts	CASIA LD	CASIA MD	CASIA HD
forehead	46	37	26
cheeks	23	17	29
mouth+upper-mouth	23	30	36
nose	26	33	36
eyes+eyebrows	46	36	26
chin	26	27	39
whole face	**23**	**17**	**16**
component-based	-	-	20

2.2.4 Support Vector Machine for classification purpose

Fake face detection is assimilated as a binary classification problem. Multiple supervised learning techniques have been investigated in the literature including χ^2 classifier (nearest neighbours with χ^2 distance), Partial least square (PLS-DA) discriminant analysis, Linear discriminant analysis (LDA) and the most popular Support Vector Machines (SVMs). In this work, we choose the SVM classifier as it has demonstrated state of the art performance in recent works. The concepts behind SVM classification and its use for the problem of fake face detection are detailed in the next section.

2.2.4.1 Support vector machines (SVM)

The SVM binary classification algorithm searches for an optimal hyperplane that separates the labelled training samples into two classes. Let $(x_i)_{i=1:n}$ be the set of training samples and $(y_i)_{i=1:n}$ their corresponding labels ($y_i \in -1, 1$), the goal of SVM is to learn a decision function (or scoring function) that maps any input sample x_i to a real value $\hat{y_i}$ close to its original label y_i. This function corresponds to the projection of the input sample x_i on the orthogonal vector to the optimal hyperplane parametrized by its direction w and offset from the origin b:

$$\hat{y_i} = w^t x_i + b \tag{2.2}$$

For separable classes, the optimal hyperplane maximizes a margin (space that does not contain any observations) surrounding itself, which creates boundaries for the positive and negative classes as illustrated in figure 2.10. The SVM margin is equal to $2/\|w\|$ and the objective is to minimize the following optimization problem:

$$\begin{aligned} \underset{w,b}{\text{minimize}} \quad & \frac{1}{2} w^t w \\ \text{subject to} \quad & y_t(w^t x_t + b) \geq 1 \end{aligned} \tag{2.3}$$

For inseparable classes, the objective is the same, but the algorithm imposes a penalty on the length of the margin (soft margin) using a slack variable ξ for every observation that is on the wrong side of its class boundary. Besides, the dual form of this optimization problem is preferred as it benefits from the kernel trick for non-linearly separable data. The kernel trick is employed to

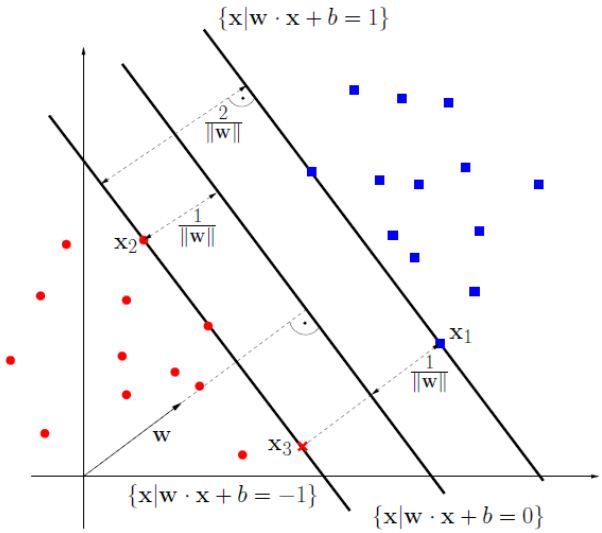


Figure 2.10: Large margin separating hyperplane for linearly separable positive class (blue) and negative class (red).

map the input data to a higher dimensional space where it becomes linearly separable without an explicit expression for the mapping (only dot products). The dual problem is formulated as:

$$\begin{aligned} &\underset{\alpha}{\text{maximize}} \quad \sum_{t=1}^{n} \alpha_i - \frac{1}{2} \sum_{k,l=1}^{n} \alpha_k \alpha_j y_k y_l K(x_k.x_l) \\ &\text{subject to} \quad 0 \leq \alpha_i \leq C \text{ and } \sum_{i=1}^{n} \alpha_i y_i = 0 \ , i = 1, .., n \text{ and KKT complementarity conditions.} \end{aligned} \tag{2.4}$$

where K is the kernel function and C is the cost parameter associated to the soft margin penalty term. The Karush-Khun-Tucker complementary (KKT) conditions implies that Lagrange multipliers α can be non-zero only for training points on the margin (Support Vectors). Greater C leads to fewer support vectors and can reduce over-fitting but at the same time can degrade the performance. Quadratic programming solvers are used to solve this dual problem. Most commonly used kernels are the following:

- linear kernel: $K(x_i, x_j) = x_i^t x_j$.
- Radial Basis kernel (RBF): $K(x_i, x_j, \sigma) = exp(-(x_i - x_j)^t (x_i - x_j)/\sigma^2)$.
- Polynomial kernel: $K(x_i, x_j, p) = (1 + x_i' x_j)^p$.
- Histogram intersection kernel: $K(x_i, x_j) = \sum_{i=1}^{k} min(x_i(k), x_j(k))$.

2.2.4.2 Feature scaling and model selection

Hsu and al. insist on the importance of scaling the features into a fixed dynamic range in [Hsu03] and recommend linearly scaling each attribute in the range of [-1,1] or [0,1] when using the libSVM package [Chang11]. The advantage of feature scaling is to avoid that attributes with large values dominate those in smaller range and also avoid numerical difficulties when computing kernel dot products. Scaling parameters are determined from training data attributes only to avoid prediction

bias. The Matlab SVM implementation of SVM prescribes feature standardization so that training attributes have zero mean and unit variance. The proposed standardization procedure takes into account the class distribution in the training set by weighting each sample with their class prior probability to handle unbalanced classification data. It is also common to perform Principal component analysis (PCA) to reduce the feature dimension so that 98% of the variance of the training data is retained.

Furthermore, SVM offers multiple kernels for mapping low dimensional data to higher dimensional space for non linearly separable data. The penalty cost C and kernel parameters are tuned on a development dataset when available otherwise a cross-validation procedure is performed on the training set. The most common parameter tuning technique uses the grid search algorithm where an exhaustive list of C and kernel parameters values are tested and those who achieve the highest classification accuracy are retained. The final model is learned using those optimal values.

2.2.4.3 SVM design for fake face LBP-based detection

We investigate the impact of feature normalization techniques and model selection on the classification stage for the LBP case study. The face region is extracted and geometrically normalized to maintain a 54 pixels inter pupillary distance. The following classification configurations are tested:

- configuration 1: LBP histogram is normalized to sum to one and a linear kernel is considered.
- configuration 2: LBP histogram is normalized to sum to one and a RBF kernel is considered.
- configuration 3: LBP histogram is normalized to sum to one and a histogram intersection kernel is considered.
- configuration 4: features are rescaled between [-1,1] and a RBF kernel is considered.
- configuration 5: features are standardized (0 mean and unit variance) and a RBF kernel is considered (Matlab default).
- configuration 6: features are standardized (0 mean and unit variance) and PCA is applied so that 98% of the variance is retained, the RBF kernel is used.

Experiments are conducted on CASIA and ReplayAttack databases. Table 2.8 reports the classification results in terms of EER for CASIA samples and HTER for ReplayAttack samples. No particular configuration stands out and the best configuration depends on the considered dataset. Feature scaling and model selection impact greatly the performance of the LBP countermeasures as performance can vary up to 100% depending on the considered data. Note that PCA normalization obtains the best EER on low quality samples and can prove to be efficient on larger features. Nonetheless, we select as best configuration the default Matlab configuration as it achieves the best results overall.

Table 2.8: Performance results (in %) for different feature normalizations and kernels.

	CASIA LD (EER in %)	CASIA MD (EER in %)	CASIA HD (EER in %)	ReplayAttack (HTER in %)
config 1	30	16	22	21
config 2	26	10	29	25
config 3	30	16	19	23
config 4	26	10	29	25
config 5	33	16	16	18
config 6	23	23	23	19

2.2.5 Comparison with multi-frame evaluation frameworks

We compare our LBP countermeasure design with the implementations of Freitas [Freitas Pereira13] and Komulainen [Komulainen13]. Results are reported in table 2.9. The proposed face registration procedure is not competitive with existing LBP countermeasure designs. It appears that averaging LBP features from multiple frames is essential and aligning faces does not compensate the need for multi-frame computation and may generate interpolation noise. Also, we believe that the traditional face detection procedure is sensitive to possible face size bias between real and fake face which generates different bounding box sizes around the face and therefore transfers this disparity to the feature extraction stage. Besides, when training the system, the missing fake faces (wrongly detected, ie section 2.2.1.3) can have a negative impact on the SVM classifier leading to worse performance.

Consequently, we adopt a simpler face extraction procedure similarly to Komulainen and al. [Komulainen13] but using the Matlab face detector based on ENCARA 2 [Castrillón07] to compete with other LBP-based countermeasure implementations. The cropping window is adjusted to contain the whole face while limiting the amount of background. All faces are detected and we display some cropped face examples resulting from this alternative face extraction procedure in figure 2.11. A multi-frame procedure is employed where a feature vector per video is obtained by averaging features from the first 2 seconds of video (50 frames at 25 fps). Although better results are obtained using more frames, we limit the time required for authentication to 2 seconds for practicability. Despite changing the face extraction procedure, the optimal face geometric normalization and classification settings holds so the whole face region is normalized to 150*135 pixels (corresponds to IPD = 54) and features are standardized before classification with an SVM-RBF classifier. With this evaluation framework, decent detection is obtained on ReplayAttack and on high quality acquisitions from CASIA database.

Figure 2.11: Exemplars of extracted faces from one client of the CASIA database.

Table 2.9: LBP performance comparisons on CASIA and ReplayAttack databases under the proposed evaluation framework.

	CASIA LD (EER in %)	CASIA MD (EER in %)	CASIA HD (EER in %)	ReplayAttack (HTER in %)
LBP(8,1,u2) [1]	-	-	-	15.16
LBP(8,1,u2) [2]	11	17	13	-
LBP(8,1,u2) [3]	4	10	0	-
LBP(8,1,u2) [4]	33	16	16	18
LBP(8,1,u2) [5]	26.6	13.9	6.67	10.6

[1] LBP results reported from [Chingovska12]. Faces are normalized to 64*64 pixels with an IPD = 33 pixels and every video frame is considered as independent samples.
[2] LBP results reported from [Freitas Pereira13]. Faces are normalized to 64*64 pixels with IPD = 33 pixels and 75 frames per video are averaged to obtain a single feature vector per video.
[3] LBP results are reported from [Komulainen13]. An extended face region is considered and no geometric normalization is performed. 50 frames are averaged to obtain a single feature vector per video.
[4] Our LBP results obtained using normalized faces with IPD = 54 pixels and only one frame is selected per video.
[5] Our LBP results obtained with the first 50 frames and using the whole face region normalized so that the inter pupillary distance is equal to 54 pixels.

2.2.6 Conclusion

Our goal is to define a processing pipeline adapted for texture-based countermeasures. First, a single frame approach relying on an advanced face registration procedure for robust face extraction is presented. This procedure has the advantage of low computational cost and extensive experiments on face geometric normalization, selection of interest regions and classification procedures could be conducted.

- Geometric normalization has a central role when examining texture information as it impacts the original quality of the recordings. Conceptually, high quality acquisitions allow texture analysis at finer scales. However, our experiments demonstrate that coarser scales provide a better texture description. Resizing face images to 54 pixels of inter pupillary distance achieves good performance while limiting the cost for computing LBP codes regardless of the sensor resolution.

- Texture from multiple face parts is analysed and we demonstrate that some parts are more or less discriminant depending on the quality of the recordings. Simply combining LBP histograms from multiple face parts does not improve the detection so the whole face is considered. Note: more complex fusion strategies based on Fisher criterion [Yang13] and [Benlamoudi15] are required to exploit correctly the different texture contributions of face parts.

- The question of feature scaling and model selection is addressed. We show that these procedures are of strong importance as performance varies significantly with the classifier configuration from one database to the other. Simple feature standardization (0 mean and unit variance) with RBF kernel is retained for all the experiments although better results can be obtained with normalization and kernel tuning on each dataset.

Comparison with existing LBP-based countermeasure designs proved that multi-frame processing provides better performance as texture variability from consecutive frames has a positive impact on fake face detection. In fact, recent studies using dynamic texture descriptors such as LBP-TOP exploit this aspect. As a result, the whole registration process is replaced by a simpler face detection method and LBP features from multiple frames are simply averaged to form more discriminant texture features. This implementation obtains competitive results compared to state of the art LBP-based countermeasure design and is retained for the rest of this study.

2.3 Evaluation of state of the art texture-based countermeasures under a unified framework

The objective of this study is to evaluate state of the art texture-based methods under the proposed framework to obtain a fair comparison of each method on the ReplayAttack, CASIA and MSU databases. We briefly present state of the art texture-descriptors employed for face anti-spoofing.

2.3.1 Description of texture descriptors

We recapitulate the different state of the art texture descriptors encountered in the literature. The comparative study will consider those state of the art texture descriptors only.

2.3.1.1 LBP and Multi-scale LBP

For clarity, the LBP formulation is briefly reminded here although already presented in section 2.1.3.2. To encode neighbourhood relationships around a given pixel p_c, N samples points p_i with $i \in \{1, N\}$ distributed evenly on a radius R around p_c are used to compute the associated lbp code as:

$$lbp_{N,R}(p_c) = \sum_{i=1}^{N} s(p_i - p_c)2^i \text{ , where } s(p) = \begin{cases} 1, & \text{if } p > 0 \\ 0, & \text{otherwise} \end{cases}$$

The normalized histogram of the lbp codes forms the traditional LBP features:

$$LBP_{N,R} = hist(lbp_{N,R})/N$$

where N is the number of pixel in the LBP image.

Non uniform patterns are grouped into a single bin of the histogram. Uniform patterns are LBP codes with at most 2 bitwise transitions 1-0 or 0-1 in their bit representation. It was shown in [Ojala02] that most of the patterns in natural images are uniform and considering only uniform

patterns adds statistical robustness while reducing significantly the number of possible patterns from 2^N to $N(N-1)+3$.

The multi-scale LBP approach of Maatta and al. [Maatta11], denoted by MS-LBP in this thesis, combines $LBP^{u2}_{8,1,3*3}$, $LBP^{u2}_{8,1}$ and $LBP^{u2}_{16,2}$ where $LBP_{N,R,b*b}$ denotes multi-block LBP features by dividing the image into $b*b$ blocks and concatenating LBP features from each block into a single feature vector.

2.3.1.2 Local binary patterns variance (LBPV)

LBPV has been introduced in [Guo10b] for texture classification tasks. It is closely related to LBP, the only difference lies in the computation of the histogram which accumulates the variance from the local region of the corresponding LBP code as follows:

$$Var_{N,R}(p_c) = \frac{1}{N}\sum_{i=1}^{N}(p_i - u)^2, \quad \text{where } u = \frac{1}{N}\sum_{i=1}^{N} p_i \tag{2.5}$$

where p_i designate the neighbouring pixels circularly distributed around p_c. Next, the computation of LBPV features is given by:

$$LBPV_{N,R}(k) = \sum_{p \in Im} Var_{N,R}(p).\delta(lbp_{N,R}(p), k), \quad \text{where } \delta(i,j) = \begin{cases} 1, & \text{if } i = j \\ 0, & \text{otherwise} \end{cases} \tag{2.6}$$

In [Kose12], Kose and al. used this scheme together with global matching to enforce rotation invariance and DoG filtering as preprocessing. They obtained satisfactory results on the NUAA database. In this article, only the simple LBPV scheme is used in order to evaluate if discriminant texture patterns have high variance like in texture classification.

2.3.1.3 Gabor features

A 2D Gabor filter consists of a sinusoidal wave modulated by a Gaussian envelope defined by:

$$G(x,y) = \frac{F^2}{\pi\gamma\eta} exp(-F^2[(\frac{x'}{\gamma})^2 + (\frac{y'}{\eta})^2]) exp(i2\pi F x') \quad \text{,with } \begin{cases} x' = xcos(\theta) + ysin(\theta) \\ y' = -xsin(\theta) + ycos(\theta) \end{cases}$$

F is the central frequency of the filter, θ is the angle between the direction of the sinusoidal wave and the x-axis of the spatial domain, γ and η are the standard deviations of the Gaussian envelope in the direction of the wave and perpendicular respectively. These last two parameters determine the shape and size of the Gaussian surface and are often referred to as smoothing parameters. The design of Gabor filter bank consists in selecting a set of filters spanning all the directions with a regular angular step at different frequencies varying with a constant ratio usually set to $\sqrt{2}$. For texture classification, the highest central frequency F_m is computed so that the half-peak magnitude iso-curve of the filter at the highest frequency touches the value of 0.5 $pixels^{-1}$ (Nyquist frequency) as suggested in [Bianconi07]:

$$F_m = \frac{\gamma}{2(\gamma + \sqrt{log(2)}/\pi}$$

According to the authors, the number of frequencies n_f and the number of orientations n_o have little effect on texture classification whereas the smoothing parameters γ and η have a significant impact. In our work, we fix $n_f = 4$ and $n_o = 6$ as in [Waris14] while smoothing parameters are chosen with grid search. Gabor features are obtained by convolving the image with each filter of the set of Gabor filters and then first and second order statistics on response coefficients are derived to form the final feature vector as:

$$GF_{\gamma,\eta} = [\mu_{11}, \sigma_{11}, \mu_{12}, \sigma_{12}, ..., \mu_{n_o n_f}, \sigma_{n_o n_f}]$$

2.3.1.4 GLCM features

Fisrt introduced by Haralick et al. in [Haralick73], gray-level co-occurrence matrices (GLCMs) are widely used in texture analysis. Considering a pairing rule between pixels, GLCM counts the number of different combinations of gray levels occurring for all pair of pixels in the image. Given a quantization factor q and a relation operator parametrized by the distance d and the direction $u \in \{\rightarrow, \searrow, \nearrow, \uparrow\}$, glcm is a q*q matrix where glcm(i,j) is the probability of having a pair of gray levels $(i, j) \in [1, .., q]^2$ for pixels satisfying the relation operator. Second order statistics are computed from this matrix to form the feature vector. In this work, we use 19 texture features as defined in [Haralick73, Soh99, Clausi02] corresponding to: Autocorrelation, Cluster Prominence, Cluster ShadeContrast, Correlation , Difference entropy, Difference variance, Dissimilarity, Energy , Entropy, Information measure of correlation1 , Information measure of correlation2 , Inverse difference (Homogeneity in matlab) , Maximum probability , Sum average, Sum entropy , Sum of squares (variance), Sum variance. The Matlab implementation of these features is provided by Patrik Brynolfsson [2].

2.3.2 Experimental protocol

Faces are extracted using the ENCARA2 [Castrillón07] face detector. Then, cropped faces are normalized to 150*135 pixels to maintain an interocular distance of 54 pixels. Features from the first 50 frames (2 seconds of video) are accumulated to form texture features for each video. Features are then standardized before classification based on an SVM classifier with RBF kernel.

Grid search is used to determine the best parameters for each descriptor on each database. To limit the computation time, we only use one frame during the grid search procedure. Also, some optimal parameters can vary between different databases so we try to selected the same parameters for different databases when performance is close to the optimum for simplicity. The list of selected parameters are reported in table 2.10.

Table 2.10: List of parameters

descriptor	parameters	CASIA	ReplayAttack	MSU
LBP,LBPV	(N, R)	(8,1)	(8,1)	(8,1)
Gabor	(γ, η)	(4,4)	(4,4)	(6,6)
GLCM	(d, q)	(1,32)	(1,32)	(1,32)

[2] https://fr.mathworks.com/matlabcentral/fileexchange/55034-glcmfeatures-glcm-/content/GLCMFeatures.m

The evaluation follows the public protocols associated with each database as described in section 1.2.3. EER is reported for both MSU and CASIA databases whereas HTER is reported for the ReplayAttack database. The countermeasure is trained on each data subset defined by the evaluation protocol.

2.3.3 Results on CASIA database

The performance of the aforementioned descriptors on the CASIA database is reported in table 2.11 in terms of EER. The best overall performance is achieved by the MS-LBP features. The traditional LBP descriptor is slightly behind in terms of overall EER and both methods clearly outperform LBPV, Gabor and GLCM based countermeasures.

Impact of sensor quality Acquisitions from three different sensors are available in the CASIA database which allows us to evaluate the robustness of countermeasures over multiple acquisition devices. Except the LBPV based countermeasure, good detection is achieved on high quality acquisitions with an EER between 6.6% and 10%. For low quality and normal quality recordings, MS-LBP features clearly outperform other countermeasures and achieve 10% and 9.4% respectively. A significant drop in performance is observed for normal and lower quality acquisitions for the other state of the art texture descriptors which highlights the superiority of MS-LBP for this database. Different texture patterns are discriminant for various sensors and MS-LBP is able to capture them. However, learning a general classifier which is able to discriminate features from multiple sensors at once is more difficult and performance decreases to 14.2% like with traditional LBP countermeasures.

Impact of attack scenarios Robustness to different spoofing attack scenarios is also investigated with separate experiments on warped, cut and video attacks. Recordings from the three sensors are considered in the evaluation as described in the public protocol. The MS-LBP descriptor outperforms other descriptors for each of these attack scenarios, particularly against iPad video attacks with almost perfect detection ($EER = 5.5\%$).

Table 2.11: EER (in %) results on CASIA database.

	low quality	normal quality	high quality	warped	cut	video	overall
LBP	26.6	13.8	**6.6**	14.4	18.8	14.4	14.4
MS-LBP	**10**	**9.4**	7.2	**10**	**12.2**	**5.5**	**14.2**
LBPV	22.7	19.4	23.3	24.4	23.3	21.1	24.6
Gabor	19.4	33.3	7.2	23.3	20	18.8	18.8
GLCM	16.1	23.3	10	17.7	17.7	25	25

2.3.4 Results on ReplayAttack database

The results on the ReplayAttack database are reported in table 2.12 in terms of HTER. Once again, the best overall performance is achieved by the MS-LBP features but decent detection is also obtained using LBP, LBPV, Gabor and GLCM features. Similarly to CASIA results, video attacks

are easier to detect compared to photo attacks as HTER reaches 11.4% and 4.6% respectively. In this experiment, the impact of the quality of the spoofing attacks is investigated using three different devices to perform the attacks going from the use of simple printed photos to digital photos displayed on iPhone and iPad. Experiments on the print attacks and iPhone yield almost perfect detection with $HTER = 3.1\%$ and $HTER = 4\%$ respectively. However, significant performance decrease is observed on iPad attacks as the screen size is big enough to render a good fake face relatively to the sensor quality.

Table 2.12: HTER (in %) results on ReplayAttack database.

	print	mobile	iPad	photo	video	overall
LBP	4.3	3.4	20.9	16.0	6.8	10.6
MS-LBP	**3.1**	4	**12.8**	**11.4**	**4.6**	**8.7**
LBPV	13.7	4.7	**12.8**	12.5	6.2	9.0
Gabor	6.8	**3.1**	17.1	**11.4**	10	12.2
GLCM	5.6	5.3	15.9	12.2	12.5	12.8

2.3.5 Results on MSU database

Table 2.13 reports the performance of the texture-based countermeasures on the MSU database. Unlike the previous databases, the best overall performance is achieved using the GLCM features with $EER = 17.5\%$.

Impact of sensor quality Interestingly, GLCM features perform well for the android sensor with $EER = 10.8\%$ and fail to detect attacks from the laptop acquisitions. Reversely, MS-LBP features obtain great detection performance on laptop acquisitions with $EER = 5.8\%$ but is inefficient when a laptop sensor is employed. This proves that the type of sensor has a significant impact on the detection capabilities of state of the art texture-based countermeasures. This huge gap in performance between the two sensors can be partly explained by the face size bias between real and fake faces as laptop acquisitions tend to produce larger fake faces when performing a full view attack. Both ReplayAttack samples and MSU laptop samples have different face size distributions as shown in figure 2.6a and 2.6b and despite the geometric normalization texture differences are emphasized by this property.

Impact of attack type Similarly to ReplayAttack database, three different attack scenarios are considered using print, iPhone and iPad attacks except that only video attacks are displayed on screen. Decent detection is obtained against iPad and iPhone video attacks with $EER = 10\%$. Looking at the DET curves, we observe that better performance is achieved on iPhone attacks compared to iPad ones in accordance with the empirical fake face quality. However, poor detection of print attacks is achieved with $EER = 15\%$ only.

Table 2.13: EER results on MSU database.

	android	laptop	iPad	iPhone	print	overall
LBP	20.8	29.1	22.5	17.5	**15**	27
MS-LBP	25	**5.8**	**10**	**10**	**15**	20.4
LBPV	25.8	15	25	15	31	27
Gabor	30	15	17.5	17.5	27.5	27.5
GLCM	**10.8**	20	**10**	**10**	20	**17.5**

2.3.6 Discussion

The real challenge in face anti-spoofing is to find features that are capable to capture discriminant information that are consistent over multiple sensors and over different attack types. When a single sensor is used, the multi-scale LBP approach of [Maatta11] outperforms other state of the art methods and decent detection is obtained despite the variety of attacks, except for the android mobile sensor of MSU database. Experiments on the ReplayAttack database showed that the type of spoofing attack has a real impact on the performance and hopefully video attacks are well detected. Performance on photo attack detection varies a lot between one attack scenario to the other. Although print attacks from the ReplayAttack database are well detected, those using high quality printouts (MSU print attacks) and performed at mid-range (CASIA print attacks) are still challenging. Besides, sensor variability adds another difficulty to the problem. The above evaluations have revealed some severe limitations of current state of the art texture-based countermeasures as significant decrease in performance is observed when multiple sensors are employed. It appears that each sensor and each attack type have inconsistent texture cues and a unique discriminative model (classifier) is not able to cope with all this variability.

Then, two directions for the improvement of texture-based methods are interesting. One could investigate a multi-classification approach to deal with each type of attack separately and then infer the right decision or one could look into other ways to characterize texture. We chose to focus on the last option and investigated relevant variants of LBP for improving the detection.

2.4 Improvement of LBP countermeasures

Considering the success of LBP-based methods in the literature, we investigate LBP variants to improve the current state of the art on texture-based countermeasures. Various LBP-based schemes have been developed to best serve multiple computer vision tasks such as object detection, face recognition and texture classification but only the MS-LBP has been designed for fake face detection. A review of LBP-based countermeasures is provided in [Pietikäinen11]. Multiple strategies have been adopted to improve the performance of the traditional LBP by handling properties such as scale invariance, rotation invariance, illumination invariance, blur robustness, noise robustness. In this work, we chose to improve texture description using color and contrast information to have a texture descriptor capable of capturing the radiometric transformations occurring during the recapturing process.

2.4.1 Enhancement with contrast information: Complete Local Binary Pattern (CLBP)

LBP naturally enjoys robustness against illumination shifts in terms of intensity but it is highly sensitive to illumination direction changes. Experiments on the ReplayAttack database revealed that detection performance drops significantly when the countermeasure is trained under controlled conditions and tested under complex lightning. So, we investigate if handling contrast information is a better way to capture discriminant texture information despite increasing illumination sensitivity because texture-based methods are more likely used under controlled illumination conditions anyway.

We select the CLBP descriptor [Guo10a] which had a lot of success in texture classification tasks. The idea of CLBP is to encode both the sign and magnitude of the local difference between the center pixel and its N neighbours. The sign part corresponds to the classic LBP codes while the magnitude part $CLBP\_M$ encodes contrast information. The magnitude codes are computed as follows:

$$clbp\_m_{N,R}(p_c) = \sum_{i=1}^{N} f(p_i - p_c)2^i \quad \text{, where } f(x) = \begin{cases} 1, & \text{if } |x| > t \\ 0, & \text{otherwise} \end{cases}$$

t is set to the average of the absolute local difference value for the whole image. Additionally, the image gray level also has discriminant information for texture analysis. This information is encoded into an additional bit denoted $clbp\_c$ in $clbp\_m_{N,R}$ computed as follows:

$$clbp\_c(p_c) = \begin{cases} 1, & \text{if } |p_c| > g \\ 0, & \text{otherwise} \end{cases}$$

where g is the average grayscale value of the image. Let us denote this feature $CLBP\_MC$.

In [Guo10a], there are two ways to fuse the sign part with the magnitude part. $CLBP\_SMC$ is obtained after joint 2D histogram computation while $CLBP\_S\_MC$ is the result of the concatenation of *lbp* and *clbp_mc* histograms. In our early experiments, we observed that $CLBP^{\S}\_M$ offers the best trade off between complexity and performance. Hence, $CLBP$ refers to the concatenation of sign and magnitude LBP components in this work.

2.4.2 Enhancement with color: HSI-LBP

We propose to improve the discriminative power of texture features by using color information. Several color texture descriptors have been combined with other low level texture features in [Tronci11]. Originally developed for Image retrieval, CEED [Chatzichristofis08a], FCTH [Chatzichristofis08b], MPEG-7 descriptors, RGB and HSV histograms have been used to synthesize the visual content of images with a good computation efficiency. In [Schwartz11], Color frequency (CF) is used to add color information to the classic HOG descriptor. One drawback of these methods for face anti-spoofing is that absolute color information is used together with texture features to describe the image, although skin color or illumination color are not discriminative between real and fake faces.

In this work, we introduce a novel texture descriptor which encodes color shifts between neighbouring pixels making it invariant to skin or illumination color shifts. Inspired from [Zhu10], we derive an LBP-based color texture descriptor from the HSI color space. LBP features are computed

from all three channels and concatenated to form the HSI-LBP features. When computing LBP on the Hue channel, the local differences between the center pixel and its N neighbours is handled so that it is comprised in $[-\pi, \pi]$ in order to characterise if the difference in Hue is counter-clockwise (bit to 1) or clockwise (bit to 0). This coding is relevant as only small local color shifts occur in homogeneous regions. We tried different color spaces and HSI obtained the best results.

2.4.3 Experimental results

Experiments are conducted on CASIA ,ReplayAttack and MSU databases and results are compared to the traditional LBP countermeasure and the MS-LBP countermeasure.

2.4.3.1 Experiments on CASIA

Both CLBP and HSI-LBP achieve better results than state of the art countermeasures as shown in table 2.14. Especially, video attacks are well detected as recaptured faces appear over-exposed because of the screen direct light source. While moderate overall improvement is obtained using the CLBP features, the use of color significantly boost the detection in both single and multiple sensor cases.

Table 2.14: EER results on CASIA database.

	low quality	normal quality	high quality	warped	cut	video	overall
LBP	26.6	13.8	6.6	14.4	18.8	14.4	14.4
MS-LBP	10	9.4	7.2	10	12.2	**5.5**	14.2
CLBP	23.3	10	3.3	13.3	16.6	8.9	12
HSI-LBP	**6.1**	**3.8**	**3.3**	**7.7**	**11.1**	6.7	**6.8**

2.4.3.2 Experiments on ReplayAttack

Similarly to results on CASIA database, decent improvement is observed on digital attacks from the ReplayAttack database when contrast and color information is taken into account.

Table 2.15: HTER results on ReplayAttack database.

	print	mobile	iPad	photo	video	overall
LBP	4.3	3.4	20.9	16.0	6.8	10.6
MS-LBP	**3.1**	4	**12.8**	**11.4**	**4.6**	**8.7**
CLBP	5.6	0.9	12.8	7.5	2.8	5.3
HSI-LBP	**3.1**	**0.6**	**5.3**	**4.7**	**0.3**	**3.7**

2.4.3.3 Experiments on MSU

Detection results on MSU database are reported in table 2.16. For this database, the contrast information is not valuable for the fake face detection as worse results are obtained using the CLBP

features compared to state of the art methods. Color information yields the best performance with $EER = 15\%$ but it is still unsatisfactory for reliability.

Table 2.16: EER results on MSU database.

	android	laptop	iPad	iPhone	print	overall
LBP	20.8	29.1	22.5	17.5	**15**	27
MS-LBP	25	**5.8**	**10**	**10**	15	20.4
CLBP	30	24	15	12.5	15	25
HSI-LBP	**20**	10	**10**	12.5	**7.5**	**15**

2.4.4 Comparison with recent cue-based methods

We have proposed two variants of LBP features to better capture the differences between real and fakes faces. While CLBP obtains mitigated results, HSI-LBP features demonstrate superior anti-spoofing capabilities on all three databases. Over-exposure and faded colors, characteristic of fake faces, are well captured by the proposed method on ReplayAttack and CASIA databases. Acquisitions from the MSU databases are still challenging especially those acquired using the Google Nexus mobile.

In this Chapter, we have focused on data-driven texture-based methods focusing on the face region only. To put our work in perspective, we compare the proposed HSI-LBP features to recent cue-based methods based on image quality assessment (IQA) [Galbally14a] and image distortions analysis (IDA) [Wen15]. Both methods exploit contrast and color information as well. Table 2.17 shows that the proposed HSI-LBP descriptor is competitive against recent cue-based methods focusing on the face region only overall.

Table 2.17: EER results on MSU database.

	CASIA (H protocol)	ReplayAttack	MSU
IDA	13.3	7.4	**8.58**
IQA	5.6	15.2	-
HSI-LBP	**3.3**	**3.7**	15

2.5 Conclusion

Methods based on texture analysis are key in video anti-spoofing or mask attacks as they don't require user-cooperation during authentication nor any additional equipment. Two non exclusive categories of texture-based countermeasures are investigated in the literature. After a thorough review of state of the art countermeasures, we have highlighted a number of grey areas concerning evaluations of countermeasures based on texture only. particularly, the various evaluation frameworks and the multiple combinations of static or dynamic countermeasures make it difficult to compare texture-based methods fairly.

Hence, a unified evaluation framework that takes into account the face region only is proposed. This restriction is necessary to handle varying backgrounds and different types of attacks (photos,

videos and masks) and attack scenarios (full view attacks and mid-range ones). An exhaustive study of the LBP countermeasure is conducted for the design of this framework and the influence of the different processing stages occurring in the fake face detection pipeline is investigated. Three main points have been highlighted. First, the face region must include the whole head. Second, texture features have to be extracted on multiple frames and then averaged to increase the detection performance. Third, face geometric normalization is necessary regardless of the sensor resolution and the optimal value is around 54 inter pupillary distance (in pixels).

Then, a fair comparison of state of the art texture-based methods (data-driven) is conducted under this unified framework for the ReplayAttack, CASIA and MSU databases. Evaluated countermeasures include LBP, MS-LBP, Gabor , GLCM and LBPV features. Overall, the MS-LBP approach obtains the best detection results but performance is still unsatisfactory for industrial specifications which require $FAR < 1\%$ for a $FRR = 0.5\%$. Sensor and attack type variabilities complicate the problem of fake face detection from a texture standpoint and additional information is required to characterize differences between real and fake faces.

Consequently, two variants of LBP features are proposed where contrast and color information is taken into account. Color texture captured by the HSI-LBP features outperforms existing methods under the proposed evaluation framework and is able to compete with recent cue-based methods. Poor detection is obtained on the recent MSU database especially on mobile acquisitions.

In conclusion, texture-based countermeasures are essential for the detection of spoofing attacks and good performance is achieved on both CASIA and ReplayAttack. Further effort needs to be paid for mobile acquisitions as results on MSU are not as good as those obtained on the other two databases. Nevertheless, we were able to assess the strength and limitations of data-driven texture-based countermeasures.

Chapter 3

Motion-based countermeasures

Contents

Motion-based countermeasures have demonstrated great potential in face anti-spoofing especially against photo attacks. Multiple strategies have been investigated going from interaction-free approaches to challenge/response methods. Appealing properties make motion-based countermeasures attractive. First, low quality sensors such as standard webcams can be used efficiently with this type of methods. Second, these methods are robust to illumination changes between two authentication attempts and are scene independent. However, motion-based countermeasures suffer from a lack of robustness against attack scenarios involving simulated motion or replayed motion. As a consequence, motion-based methods have been developed to detect photo attacks primarily and complementary texture or liveness counter-measures are employed to detect video or mask attacks. We believe that the current state of the art lacks experiments on exclusive motion-based countermeasures against different replay attack scenarios. After a complete review of exclusive motion-based countermeasures in photo and video attack detection, key motion cues are discussed and we argue that replay-attacks can be efficiently detected using motion-based countermeasures

in many cases. A novel data-driven motion-based countermeasure is presented and extensive experiments are conducted on ReplayAttack, CASIA and MSU databases. The proposed method takes advantage of the Conditional Local Neural Fields (CLNF) face tracking algorithm to extract rigid and non-rigid face motions in real time. Similarly to the bag-of-words feature encoding, a vocabulary of motion sequences is constructed to derive discriminant mid-level motion features using the Fisher vector framework. Finally, we investigate the impact of camera motion on the proposed countermeasure for mobile applications.

3.1 State of the art of motion-based countermeasures

In the literature, motion-based methods are primarily designed to detect photo attacks whereas additional countermeasures based on other cues are implemented to cope with video or mask attacks. In that regard, works mentioned in this review may be incomplete as only the motion related parts are discussed. Following our general classification of anti-spoofing methods in chapter 1, motion-based countermeasures are broken down as cue-based methods and data-driven methods. A general overview of existing motion-based countermeasures is laid out in table 3.1. A detailed analysis of each method is presented below.

3.1.1 Cue-based motion countermeasures

Existing motion-cue-based methods rely essentially on three types of cues for photo attack detection: planar effect, liveness (natural movements) and face-background motion correlation. These methods are very specific to the attack scenario and authentication protocol. Planar effect is observed when out-of-plane head rotations are simulated by photo warping or photo rotations so countermeasures based on this cue need yaw and pitch head motion during authentication. Liveness-based methods fail against spoofing scenarios where eyes and mouth motions are simulated. Face-background motion correlation is relevant only when both the face and the background are fake in close-up attack scenarios. Despite these limitations, multiple methods have been proposed in the literature.

Methods based on planar effect The 3D structure of real faces can be revealed from motion. Three main ideas have been exploited for face 3D assessment: face part motion consistency, geometric invariants and structure from motion.

- **Face part motion consistency:** In [Kollreider07a], Kollreider and al. analyse the trajectories of face parts to detect photo attacks. They propose a robust face detection procedure relying on fast optical flow estimation called Optical Flow of Lines (OFL). Both vertical and horizontal motions are encoded into a complex optical flow image denoted $OF_{im} = v_x + i\ v_y$. This map is exploited for both face detection (face center) and anti-spoofing. A model-based local Gabor decomposition together with SVM experts are employed to detect left and right ears for more robustness. Region of interests (ROIs) are defined at face center and ears locations and average velocity in each ROI is computed from pixels that exhibit sufficient motion (greater than half its maximum value). Only the maximum of horizontal or vertical velocity is retained when computing the average velocities to focus on primary movements. Left and right liveness ratios (c_l and c_r) compare the velocities of the face center and those of left or right ear. Values greater than 1 indicates a real face as real faces exhibit higher velocity near

Table 3.1: Summary of motion-based anti-spoofing countermeasures. The column 'attack' corresponds to the spoofing attack scenarios handled by the proposed countermeasure. Column 'Database' mentions the name of the database used to validate the countermeasure. Column 'Protocol' indicates which type of movement is present during authentication and reflects the level of interaction required by the countermeasure.

Cue-based					
Reference	Motion cues	Methodology	Attack	Database	Protocol
2007, Pan and al. [Pan07]	liveness	Eyeblink detection	warped print	Proprietary	neutral pose
2007, Kollreider and al. [Kollreider07a]	planar effect	Comparison of face center (nose) and ears motion	warped print	Proprietary	yaw + pitch
2007, Kollreider and al. [Kollreider07b]	planar effect + liveness	Comparison of face center (nose) and ears motion + lip reading	warped print, video	Proprietary	challenge response
2008, Kollreider and al. [Kollreider07b]	planar effect + liveness	rasterflow + eyeflow	warped print, eye-cut print	Proprietary	yaw + pitch
2009, Bao and al. [Bao09]	planar effect	Detection of basic planar object motion using OFF	warped print	Proprietary	yaw + pitch
2011, Anjos and al. [Anjos11]	FBC	FBC from frame difference intensity measure	print	PrintAttack DB	neutral pose
2011, Tronci and al. [Chakka11, Tronci11] (AMILAB)	FBC + liveness	FBC estimated from foreground extraction method + eyeblink detection	print	PrintAttack DB	neutral pose
2011, Yan and al. [Chakka11, Yan12, Chingovska13a] (CASIA) (LNMIIT)	FBC + liveness	Non-rigid motion + OFR	photo	ReplayAttack DB	neutral pose
2012, De Marsico and al. [De Marsico12]	planar effect	Geometric invariants	warped print	HONDA + NUAA	yaw + pitch
2013, Wang and al. [Wang13]	planar effect	3D face structure recovery from motion	print	Proprietary	yaw
2014, Anjos and al. [Anjos14]	FBC	FBC from OF	photo	PhotoAttack DB	neutral pose
Data-driven					
2011, Lorenzo and al. [Chakka11] (SIANI)	Face parts location and difference image		print	PrintAttack DB	neutral pose
2013, Bharadwaj and al. [Bharadwaj13, Bharadwaj14]	HOOF		photo, video	ReplayAttack DB + CASIA DB	neutral pose
2013, Warris and al. [Waris14]	STACOG		photo, video	ReplayAttack DB	neutral pose

the nose compared to ears when the head follows yaw and pitch movements. Also, center and sides move in opposite directions so the sign of left and right ratios are liveness indicators. Thus, a liveness score is obtained as follows:

$$L = \frac{1}{4}((|c_r| > \tau) + (|c_l| > \tau) + (sign(c_r) < 0) + (sign(c_l) < 0)) \tag{3.1}$$

In [Kollreider08], the same authors propose another face parts motion consistency countermeasure using a face segmentation made of 5 vertical stripes instead of semantic face parts (ears and face center). Face parts detection errors are no longer an issue and the average motion (optical flow magnitude) in each stripe (rasterflow) forms a wedge-pattern ('Λ' reflecting the depth of the face) as side motion is lower than the central one for live faces.

- **Geometric invariants:** In [Bao09], Bao and al. focus on the regularity of the optical flow field to detect photograph spoofing attempts. Their intuition is that 3D object optical flow

field (OFF) is irregular and more complex compared to planar object OFF. Six quantities are computed from the OFF as features to characterize rigid planar object motion as a combination of four basic movements: translation, rotation, moving forward and moving backward. One major drawback is the illumination sensitivity and the high computing cost related to the optical flow computation. Furthermore, sufficient motion is required as Anjos and al. obtained poor results on the PhotoAttack database in their review of motion-based countermeasures in [Anjos14].

In [De Marsico12], De Marsico and al. exploit geometric invariants for detecting replay attacks through a user-cooperative authentication process. A combination of the Viola-Jones' algorithm with an Extended Active Shape Model (STASM) is used to detect facial landmarks. Six subsets of these landmarks are selected to compute 2D image geometric invariants (cross ratios). Two types of cross ratios are employed: cross ratios of four collinear points and cross ratio of five coplanar points. Those cross ratios are invariant to rotations provided that the points satisfy collinearity/coplanarity constraints. Hence, one can assess if the moving face is a 2D rigid solid object by inspecting if the cross ratios remain constant over time. Experiments showed perfect results on the combined HONDA [Lee05] and NUAA [Tan10a] databases when fast yaw and pitch movements of the head are observed. The method requires obvious movement to work well as performance drops significantly when the motion is too slow. Also, pitch motion is not enough to detect photo attacks.

- **Structure from motion:** Another method using facial landmarks is proposed by Wang and al. [Wang13]. Facial landmarks are tracked by the CLM algorithm of Saragih [Saragih10] in order to recover the sparse 3D structure of the face. Their method estimates the camera projection matrix $P = K[R, t]$ from two views where K is the camera intrinsic matrix and [R,t] is the relative pose (camera extrinsic matrix relating to rotation and translation for the second view). Then, a triangulation algorithm estimates the 3D structure of the detected face by minimizing the reprojection errors in both images with a soft constraint forcing the solution to remain close to a 3D face model acquired experimentally. Multiple views are selected to refine the estimation of the above parameters. Features are derived by aligning the recovered 3D structure upfront using the previous face model to compare the 3D face structure of faces under the same view point. An SVM classifier is then trained to distinguish the genuine and fake faces. Perfect detection is achieved on warped photo attacks regardless of the sensor used for the experiment.

These methods yield outstanding results regardless of simulated motion such as warping, translation and rotations of the spoofing photos. However, one limitation is the high-level of interaction required during authentication as sufficient out-of-plane motion such as yaw or pitch head motions are necessary.

Methods based on face-background motion correlation Face-background motion correlation approaches address the problem of close-up photo attack detection where the picture covers the whole view in order to hide the spoofing medium. The whole scene is fake so background and face regions follow the same motion (highly correlated).

- **Motion intensity correlation:** In 2011, Anjos and Marcel [Anjos11] introduced a motion-based method relying on the high correlation between background and face motion to detect photo-attacks. The motion intensity is calculated in both regions of interest (face and background) by a simple gray-scaled frame-difference. An area-based normalization is then

performed in order to compare the motion in both RoIs and check if they are decorrelated. In [Chakka11, Tronci11], Tronci and al. use foreground object detection to detect face moving pixels. The face/background motion correlation measure is obtained by dividing the number of moving face pixels over the total number of pixels in the image. In [Chakka11, Yan12, Chingovska13a], Yan and al. use a similar face/background motion consistency measure. Foreground detection is performed and motion entropy is measured from the ratio of foreground pixels over the total number of pixels. Additionally, another consistency measure is defined as the χ^2 square distance between the motion trend of the face (number of foreground pixels in the face region relative to the number of face pixels over time) and the motion trend of the background.

- **Motion orientation correlation:** In 2014, the authors improved their previous strategy by computing foreground/background motion correlation with Optical Flow [Anjos14] instead of motion intensity. Almost perfect result ($HTER = 1.5\%$) is achieved on photo attacks of the ReplayAttack database corpus.

Compared to methods relying on planar effect, methods based on face-background motion correlation apply even when limited motion is available as the user remains still with a neutral facial expression in front of the sensor during authentication. No interaction between the user and the authentication system is required assuring a user-friendly utilisation. However, this cue is very specific to close-up attack scenarios and no longer applies if only the face is fake as in mid-range attacks. Simple close-up photo attack variant which consists in superposing a background image and a cut out face picture can bypass such countermeasures.

Methods based on liveness cues Liveness cues refer to subconscious motions of a live face such as eye blink, mouth movement and expression changes. Indeed, humans naturally exhibit head and facial movements related to breathing.

- **Eye blink:** In [Pan07], Pan developed a real-time eye-blink detector for liveness detection from a generic web-cam for photograph anti-spoofing purpose. A discriminative measure of eye closure is derived from the adaptive boosting algorithm for computational efficiency and embedded into the contextual model to achieve high detection accuracy. The proposed two-eye detection method is robust as it achieves 91% accuracy with black frame glasses and 98% without them on the ZJU eye-blink database [Pan07]. Furthermore, perfect detection is achieved against photo attacks on their self-collected database despite simulated live movement by moving/bending the photograph. This method has been used in [Kollreider07a] and [Tronci11] as a complementary countermeasure. Another eye-blink measure is proposed in [Kollreider08] which computes the average magnitude of the optical flow field at eye regions and divides it by the average motion magnitude of the face region. This way, only the contribution of eye motion is measured.

- **Lips movements** In [Kollreider07b], the authors employ a lip reading detector in addition to their face parts motion consistency countermeasure to cope with video attacks. A challenge/response countermeasure checks if uttered digits match the requested digits from lip movements to cope with photo and video attacks.

- **Facial expressions** In [Yan12], non rigid motion cues are computed from the residual of batch image alignment. The ratio between eye and face non-rigid motions is used to measure the contribution of non-rigid movement at the eye regions. The proposed motion features obtain

high detection accuracy (over 90%) against print attacks from the ReplayAttack database and 81.67% against print attacks from CASIA database.

While the presence of liveness cues indicates that a real access is attempted, the absence of liveness cues only gives a high probability that a picture is being displayed as even real faces can have low vitality signs during certain authentication attempts. Although low interaction is necessary during authentication, a relatively long time is required (> 5s) to accumulate enough vitality signs to have a reliable liveness measure. Besides, these countermeasures are easily circumvented by simulated movements through cut out eye and mouth regions of the printed fake face. For that reason, existing works often employ liveness-based countermeasures as complementary measures to increase the robustness against photo attacks of their anti-spoofing countermeasure based on other cues.

3.1.2 Data-driven motion-based countermeasures

Data-driven motion-based countermeasures focus on motion analysis to reveal differences between natural face movements and fake ones. Dense features are used to characterize movements of the face and a classifier is trained to discriminate real and fake faces. As such, these methods handle multiple discriminative motion cues blindly and deal with a wider range of attack scenarios provided that some training data is available for each scenario.

Face parts locations and difference image In the first IJCB competition on 2D face anti-spoofing countermeasures [Anjos11], Lorenzo and al. (SIANI research team) proposed a method based on the face part locations and the difference image. Face, eyes, nose and mouth are detected by the ENCARA2 software and basic statistics (mean and variance) are computed from each face element position over time. These features detect possible face distortions (unnatural face part locations) and the amount of motion at each face region. In addition, a measure based on the difference image in face, non-face , eyes and mouth regions is computed to detect face appearance changes. Finally, a Bayesian Network classifier is used to distinguish genuine face from fake ones. Low performance is achieved on the Print-Attack database with $HTER = 10.63\%$ but better detection may be achieved with an interaction based authentication protocol where yaw and pitch head movement are present.

Histogram of oriented optical flow (HOOF) In [Bharadwaj13], Bharadwaj and al. use Eulerian video magnification (EVM) [Wu12] as preprocessing and Histograms of Oriented Optical Flow features (HOOF) [Chaudhry09] to detect micro-facial expressions. Their method densely describes the motion of each face local block as a time series and classification is performed using linear discriminant analysis (LDA) to handle high dimensional data. They achieved state of the art results on the "Print attack" and "Replay attack" database with almost perfect detection. One key aspect of this approach is that feature dimension depends on the length of the acquisition. To avoid constraint on video length, the authors proposed to segment the video into videolets (short sequences of 25 frames) and aggregate the classification scores obtained on each videolet [Bharadwaj14]. Perfect detection is achieved on ReplayAttack but low performance is obtained on CASIA database. Surprisingly, motion magnification preprocessing has no impact on the HOOF countermeasure.

Space-time auto-correlation of gradients (STACOG) In [Waris14], the authors also employ motion magnification as preprocessing. They use space-time auto-correlation of gradients (STACOG) [Kobayashi12] features computed on the whole scene to characterize movements. Frame-based features are extracted from spatio-temporal volume of D time duration every N frames. Features are then fed to a discriminant classifier to detect photo and video attacks and manage $HTER = 9.12\%$ on the ReplayAttack database using only half the training data. This method is reminiscent of LBP-TOP countermeasure as both dynamic and static components are captured. Nonetheless, we mention this work in this study as it was originally introduced for motion recognition purposes in [Kobayashi12] and as a motion-based countermeasure in [Waris14].

3.1.3 Overview

Both data-driven and cue-based methods have demonstrated great results for photo attack detection. Countermeasures based on planar effect are robust to any photo attack scenarios at the expense of high interaction-based authentication. Methods based on face-background motion correlation and liveness cues lack robustness against attack scenarios with simulated or replayed motion. While limitations of existing cue-based methods are clear, understanding the success or failure of data-driven methods is more challenging. Close-range video attacks from the ReplayAttack database can be well detected using data-driven motion based-methods [Bharadwaj13] but poor detection results are obtained on mid-range video attacks from CASIA-FASD. To the best of our knowledge, no convincing explanation has been given to explain this disparity and more generally we believe that the current state of the art lacks a deep analysis of the strengths and limitations of recent motion-based countermeasures since they are usually combined with texture-based countermeasures and more effort is dedicated to the assessment of the whole method.

For this reason, we propose a novel motion-based countermeasure halfway between data-driven and cue-based approaches which exploits natural and unnatural motion cues using data-driven techniques without any assumption on the attack type and without any cooperation of the user during authentication. Compared to fully data-driven methods relying on generic motion descriptors, the proposed method takes advantage of the fact that the face is a 3D deformable object to derive rigid and non-rigid facial movements allowing an easier interpretation of the results. This motion characterization enables the detection of:

- Natural motion cues: rigid movements associated with head rotations and translations and non-rigid facial movements related to expression changes.
- Unnatural motion cues: rigid and non-rigid unnatural motion due to simulated movements or hand-shaking movements.

3.2 Description of a new motion-based countermeasure

Distinct face movements (macro-movements) of real faces are composed of facial action units (FACS) [P. Ekman78] and natural head movements such as yaw and pitch and subtle movements due to respiration. Reversely, typical fake face movements are shaking motion, translation or warping of the face region (for warped print attacks) and any other unnatural movements generated when performing the attack. As it is difficult to identify clearly which movement is natural or not, a

data-driven approach is adopted to learn the distribution of fake and real face movements from motion features derived from the automatic behaviour analysis framework. Constrained Local Model (CLM) face tracking algorithms are now capable of facial landmark detection, pose estimation and facial action unit recognition [Baltru16] in addition to face extraction as illustrated in figure 3.1. Originally designed for automatic facial behaviour analysis, these extra features are also helpful for anti-spoofing purposes. The proposed approach is halfway between face parts motion analysis and dense optical flow analysis which offers a better trade-off between efficiency and complexity compared to existing motion-based countermeasures.

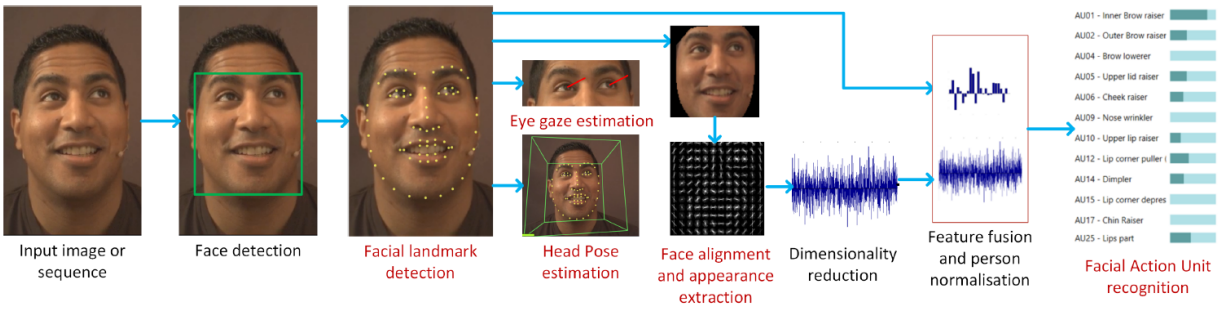


Figure 3.1: OpenFace behaviour analysis pipeline [Baltru16].

The proposed method takes advantage of the face detection based on the Constrained Local Neural Fields framework of Baltrusaitis [Baltrusaitis13] to jointly extract the face and its shape characteristics over time. Rigid and non-rigid shape parameters are computed in real time for each frame. Rigid and non-rigid low level motion parameters are derived by simple time derivation. At this point, these low level motion parameters form a two dimensional signal where the first dimension corresponds to the number of shape parameters and the second dimension is the time axis. A classical way to handle this type of signal for classification tasks is to map the low-level motion parameters into a more discriminative high-level representation using a codebook. In our case, words represent the underlying micro-movements at the foundation of recognizable face movements (such as hand-shaking motion). They are made of short sequences of low-level motion parameters either selected randomly among training data or derived from a clustering procedure. We adopt the Fisher kernel framework to derive discriminant mid-level motion features from short sequences of motion parameters (low-level motion features) which are then fed to a linear SVM classifier with default parameters. The full pipeline of the proposed countermeasure is illustrated in figure 3.2.

3.2.1 Extraction of low level motion features

Low-level motion features extraction consists in two steps as illustrated in figure 3.2 (blue box). First, facial landmark detection and tracking are performed using the CLNF algorithm from the OpenFace toolkit publicly available[1]. Then, motion sequences are computed throughout the whole video.

Shape parameters extraction Rigid and non-rigid shape parameters $\boldsymbol{p}$ are computed for each frame. They result from the Constrained Local Neural Fields (CLNF) landmark detection framework. Shape parameters control the transformations required to fit the observed face shape

[1] https://github.com/TadasBaltrusaitis/OpenFace/wiki

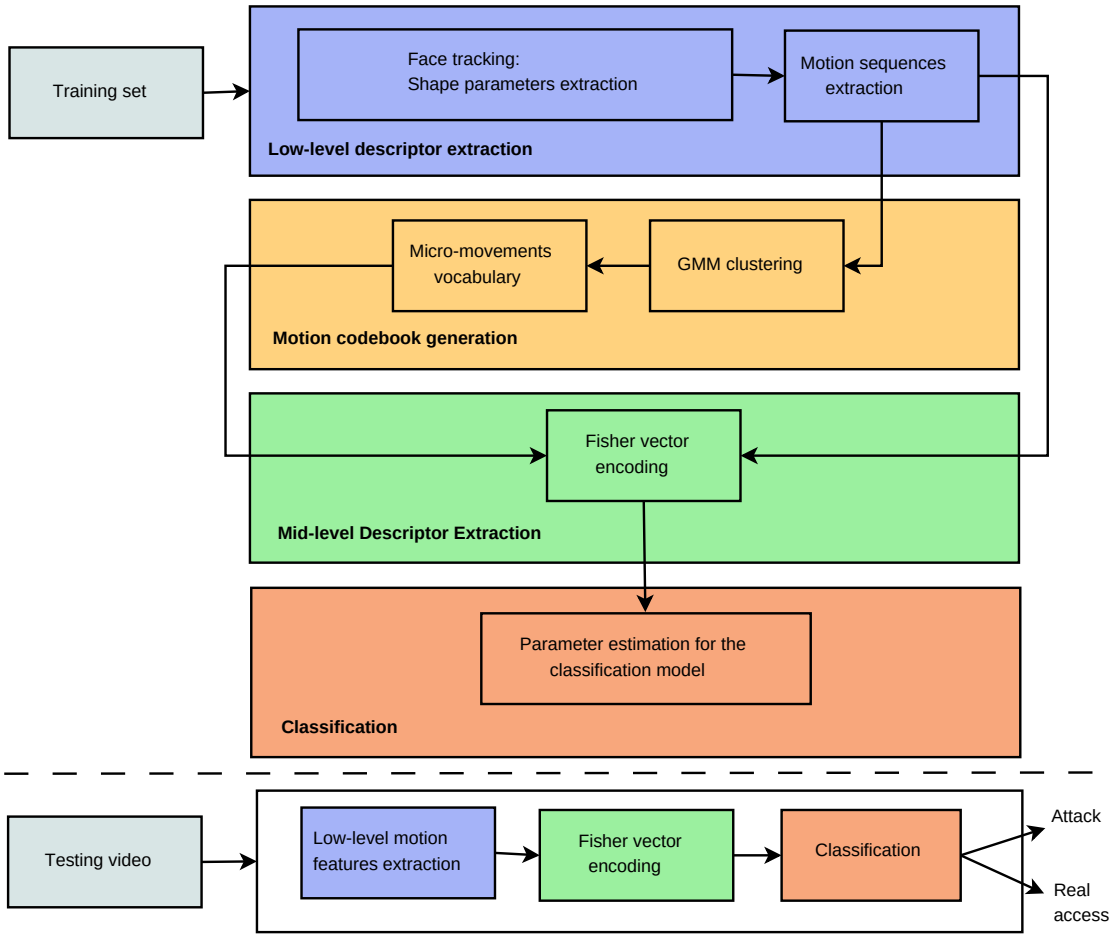


Figure 3.2: Diagram of the proposed method.

$X = \{x_i\}_{i=1:68}$ (set of 68 facial landmarks) to a reference shape $\bar{X} = \{\bar{x}_i\}_{i=1:68}$ (corresponding to the average set of facial landmarks for multiple identities) via a point distribution model (PDM). They are made of 6 rigid shape parameters including a scaling term s, translation $\boldsymbol{t} = [tx, ty]$ and rotation $\boldsymbol{\theta} = [\theta x, \theta y, \theta z]$ terms and 34 non-rigid parameters $\boldsymbol{q}$ controlling local shape deformations. The 2D location of the i^{th} landmark $\boldsymbol{x_i} = [x_i, y_i]^t$ is controlled by the parameters $\boldsymbol{p} = [s, \boldsymbol{t}, \boldsymbol{\theta}, \boldsymbol{q}]$ through the PDM:

$$\boldsymbol{x_i} = s.R_{2D}.(\bar{\boldsymbol{x}}_{\boldsymbol{i}} + \boldsymbol{\Phi_i q}) + \boldsymbol{t} \tag{3.2}$$

where R_{2D} corresponds to the first two rows of the rotation matrix associated with Euler angles $\boldsymbol{\theta}$, $\bar{\boldsymbol{x}}_{\boldsymbol{i}} = [\bar{x}_i, \bar{y}_i, \bar{z}_i]$ is the mean value of the i^{th} landmark 3D location and $\boldsymbol{\Phi_i}$ is a 3*34 principal component matrix. For details on the estimation of $\boldsymbol{p}$, please refer to the original publications of Baltrusaitis [Baltru16, Baltrusaitis13].

Motion sequences extraction From these shape parameters, rigid and non-rigid motion parameters are directly computed by using the first temporal derivative of $\boldsymbol{p}$. Then, short sequences of N frames are extracted with maximum overlapping between successive sequences ($N-1$ frames in common) to form the final low-level motion features. The selection of N is discussed in section 3.3.1.

3.2.2 Fisher kernel framework

The second step is inspired by the work of Perronin and al. [Perronnin07] in image categorization, but the Fisher kernel framework is applied on motion sequences instead of image patches. Fisher vectors can be seen as a generalization of Bag-of-Visterms (BOV) which map a given input (fixed dimensionality but variable length) into a fixed size feature vector which can then be fed to a linear classifier for classification tasks. Extending the description of the word distribution over the input from simple word count (0 order statistic) to higher order statistics (first and second) enable a more complete embedding and the use of a more compact dictionary. Fisher kernel classification framework has three stages: codebook generation, Fisher vector encoding of the input signals and classification.

- **Codebook generation**: A Gaussian Mixture Model (GMM) composed of K Gaussians is learned to model the distribution of motion sequences in a video regardless of class information. The GMM parameters $\Theta = \{\mu_k, \Sigma_k, \pi_k\}$ are estimated using the Expectation Maximization (EM) algorithm with random initialization (select K data points at random for each mode means, initialize the individual covariances as the covariance of the data, and assign equal prior probabilities to the modes). The selection of K is discussed in section 3.3.1.

- The actual encoding corresponds to the improved version of Fisher vectors (IFV) proposed in [Perronnin10]. Let $X = (x_1, ..., x_T) \in \Re^{D \times T}$ be the set of T overlapping motion sequences extracted from a given input video. Let q_{tk} denote the posterior probability of the sequence x_t to belong to class k, the normalized gradient of the log-likelihood with respect to the mean and covariance parameters is given by vectors $\boldsymbol{u_k}$ and $\boldsymbol{v_k}$ respectively:

$$q_{tk} = \frac{\exp[-0.5(x_t - \mu_k)^t \Sigma_k^{-1}(x_t - \mu_k)]}{\sum_{k=1}^{K} \exp[-0.5(x_t - \mu_k)^t \Sigma_k^{-1}(x_t - \mu_k)]}$$

$$\boldsymbol{u_k} = \frac{1}{T\sqrt{\pi_k}} \sum_{t=1}^{T} q_{tk} \Sigma_k^{-1}(x_t - \mu_k)$$

$$\boldsymbol{v_k} = \frac{1}{T\sqrt{2\pi_k}} \sum_{t=1}^{T} q_{tk}[(\Sigma_k^{-1}(x_t - \mu_k))^2 - 1]$$

  The IFV is the concatenation of all D dimensional vectors $\boldsymbol{u_k}$ and $\boldsymbol{v_k}$ for $k = 1, .., K$ leading to a 2*D*K feature vector. Signed square rooting is employed to obtain a less sparse FV which is better handled by dot-product or L2 distance generally used by the linear classifier. Also, L2 normalization is performed to remove the dependence of the proportion of (class) specific and independent (natural) motion.

- Classification is performed using a linear SVM classifier with C = 1.

The derivation of Fisher vectors supposes the independence of each sequence x_t which is not satisfied in our case as overlapping sequences are used. We show that even without this assumption, IFV still provides a discriminant representation for anti-spoofing tasks.

3.3 Experimental setup

In this section, we discuss parameter tuning and implementation choices for the proposed countermeasure. First, dictionary size and sequence duration are discussed. Next, the fusion strategy

of rigid and non-rigid motion cues is presented. Then, dictionary construction is investigated. Finally, the influence of video duration on the discriminative power of the improved Fisher vector is discussed.

3.3.1 Parameters selection

The choices of the sequence of motion duration N and the number of dictionary elements K are crucial to obtain discriminative mid-level features representation. As both rigid and non-rigid low-level motions are different by nature, parameters are tuned with respect to each type of motion. We select the best parameters for all three databases (ReplayAttack, CASIA and MSU) using a grid search strategy with $N \in \{1, 5, 10, 15, 20\}$ and $K \in \{20, 30, 40, 50, 60\}$. In this experiment, the motion vocabulary is learned in an unsupervised manner on the training set for each database as suggested in [Perronnin07].

- **Influence of vocabulary size:** To isolate the influence of K on the detection, performance results are marginalized with respect to N. Figure 3.3 plots the performance of IFV based on rigid and non-rigid motions for each database as a function of the vocabulary size. The HTER is reported for the ReplayAttack database whereas the EER is used for MSU and CASIA databases. The dictionary size has little influence on the overall performance of the IFV. The number of atoms is fixed to 50 for both rigid and non-rigid motion cues to ensure a good trade off between performance and complexity.

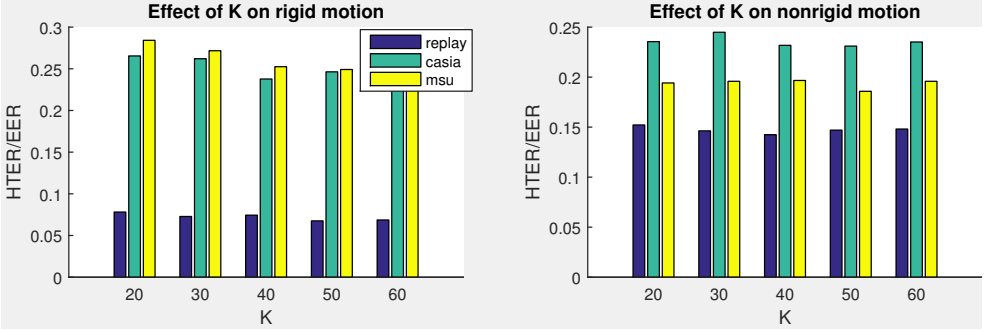


Figure 3.3: Influence of K

- **Influence of time window:** To isolate the influence of N on the detection, results are marginalized with respect to K. Figure 3.4 plots the performance of IFV based on rigid and non-rigid motions for each database as a function of the sequence length. The time window has a significant impact on the discriminative power of rigid motion features. Short sequences of 5 frames obtain the best results for CASIA and ReplayAttack databases whereas 10 frames are the best for MSU database arguably because videos are captured at a higher frame rate (30 fps for MSU compared to 25 fps for CASIA and ReplayAttack).

  The time window has little effect on non-rigid motion for ReplayAttack and MSU databases as the maximum and minimum HTERs (EER) do not exceed 4% whereas it has a significant impact for CASIA database. This observation reflects the specificity of CASIA database which includes frowning, talking and smiling motions in real accesses and print attacks involve photo bending and warping, leading to more complex non-rigid motions. In comparison, both MSU and ReplayAttack databases contain limited non-rigid motions such as slight pose variations and eye-blinks.

In the end, the sequence length of rigid and non-rigid motion features are fixed to 5 frames for both ReplayAttack and CASIA databases while 10 frames are used for MSU database.

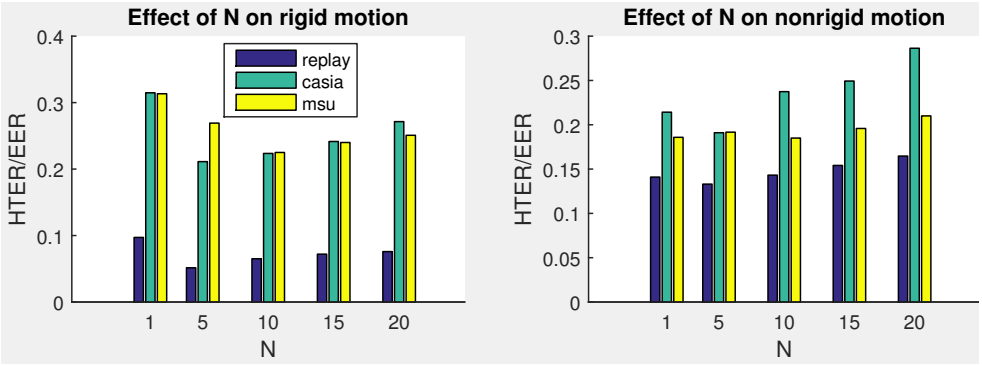


Figure 3.4: Influence of N (in frame number)

3.3.2 Fusion of rigid and non-rigid motion cues

To take advantage of both rigid and non-rigid motion cues, multiple fusion strategies are investigated. As selected sequence duration is identical for both types of motion, the first option is to concatenate both rigid and non-rigid motion sequences and to learn a single dictionary to compute the improved Fisher vectors. The second option consists in concatenating the Fisher vectors based on rigid motion and those based on non-rigid motion (mid-level) before classification. The last option combines both motion cues at score level using the average score.

Table 3.2: Fusion of rigid and non-rigid cues

HTER / EER (%)	rigid	non-rigid	Fusion		
			low-level	mid-level	score-level
ReplayAttack	4.6	13.7	5.0	3.6	4.5
CASIA	20	17	18.7	19.1	19.1
MSU	20	19	10	12	20

Table 3.2 reports classification results for each fusion scheme on ReplayAttack, CASIA and MSU databases. First, we observe that feature fusion obtains better results compared to single features except for the CASIA database where non-rigid motion features yield $EER = 17\%$ and the best fusion result is $EER = 18.7\%$. Slight improvement is obtained on the ReplayAttack database using the mid-level fusion strategy as HTER drops from $HTER = 4.6\%$ when rigid motion features only are used to $HTER = 3.6\%$. Greater improvement is achieved on the MSU database using low-level (or mid-level fusion) as EER drops from 19% to 10% (or 12% respectively). The best fusion scheme is different from one database to the other as the discriminative power of rigid and non-rigid motion features also vary. For simplicity, we choose low-level feature fusion for all the experiments because this method obtains the best performance on CASIA and MSU and it is very simple.

3.3.3 Design of the Motion Vocabulary

In [Perronnin07], the authors demonstrate that Fisher vectors trained in an unsupervised manner (class unaware) achieve similar performance as those obtained from a supervised approach in image classification. To check this property for motion-based anti-spoofing, a simple supervised vocabulary construction is designed by learning one vocabulary per category. In our experiment, the categories correspond to real accesses, photo attacks and video attacks. Class specific vocabularies are then concatenated to form the final dictionary used to derive the Fisher vectors. Table 3.3 compares both supervised and unsupervised approaches. Little improvement is observed on ReplayAttack database as HTER decreases to 4.4% whereas worse results are achieved on CASIA and MSU databases. Additionally, the dictionary size (and feature size) for the supervised approach is three times bigger than the unsupervised one leading to longer computation time. Therefore, the unsupervised approach is retained.

Table 3.3: Class aware dictionary learning

HTER / EER (%)	unsupervised	supervised
ReplayAttack	5	4.4
CASIA	18.7	20.2
MSU	10	15.4

3.3.4 Minimum video duration

Face recognition system specifications limit the duration of authentication as a matter of practicality. In this context, we investigate the minimum authentication duration to obtain decent anti-spoofing detection with the proposed countermeasure. Experiments on ReplayAttack, CASIA and MSU databases are conducted with increasing duration for each authentication attempt. Results are reported in figure 3.5. Non-rigid motion features require at least 4 seconds of accumulation to reach the minimum error rate whereas a longer period is necessary for rigid motion features. Although better results can be achieved by imposing longer authentication duration, the proposed countermeasure still yields decent performance when limited authentication duration is specified ($\approx$ 2 seconds is usually the time required for authentication) and can be employed in real world face recognition systems.

3.3.5 Discussion

In this section, we have determined the best time window (N = 10 frames for MSU-MFSD, N = 5 frames for ReplayAttack-DB and CASIA-FASD) and vocabulary size (K = 50) to capture both rigid and non-rigid discriminant movements. Experiments over various fusion strategies showed that low-level feature fusion performs well and is selected for the rest of this study. Also, we have evaluated the influence of the authentication duration on the proposed method and demonstrated decent results in real world situations. For the rest of this study however, the full video is retained to compute the improved Fisher vectors for comparison with state of the art motion-based countermeasures. Experiments in the next sections are conducted using this optimal configuration.

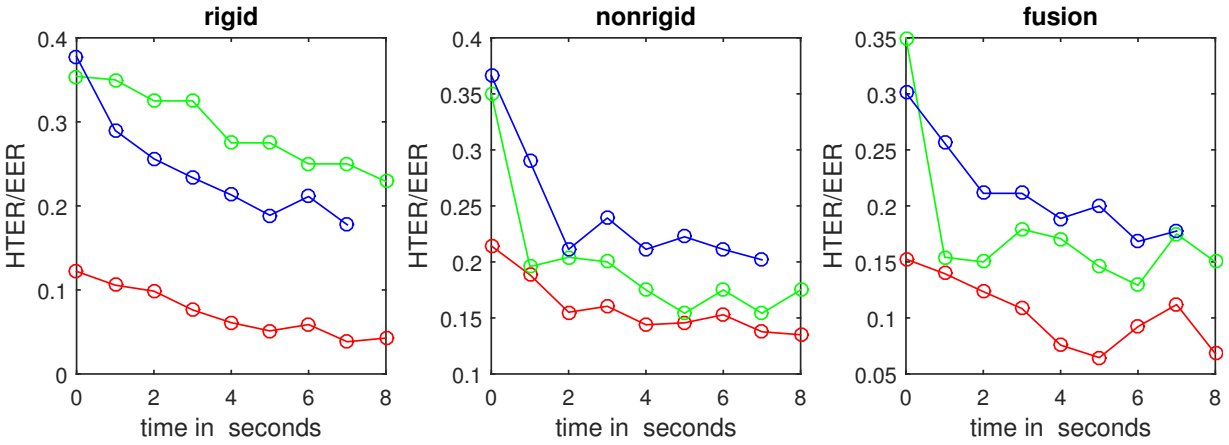


Figure 3.5: Performance of the proposed countermeasure in function of the authentication duration. The red curve corresponds to the HTER measured on the ReplayAttack database. The blue and green curves correspond to the EER measured on the CASIA and MSU databases.

3.4 Experimental validation

In this section, experiments are conducted on ReplayAttack, MSU and CASIA databases to cover a large variety of attack scenarios. A detailed evaluation of the proposed countermeasure is conducted on photo attacks and video attacks. Then, we compare our method with state of the art motion-based countermeasures. Finally, we discuss the robustness of the proposed method with respect to the sensor choice.

3.4.1 Evaluation of the proposed countermeasure against photo attacks

In this section we investigate the contribution of both rigid and non-rigid motion cues to detect photo attacks when no interaction during the authentication phase is enforced. Photo attacks include printed and digital face pictures. Two different scenarios of photo attacks are distinguished in this study as they raise different challenges from a motion standpoint: (i) still photo attacks and (ii) photo attacks with simulated motion. The first case comprises photo attacks displayed at close-range in order to hide the borders of the spoofing medium contained in ReplayAttack and MSU databases. Because background is also part of the spoofing, simulating real face motion is more difficult and the considered scenarios only hold still the photo in front of the sensor either by hand or using a fixed support. The second case refers to mid-range photo attacks from the CASIA database where only the face region is printed allowing impostors to simulate liveness by moving around, warping or cutting out the eye regions of the printed face.

3.4.1.1 Still photo attacks

The easiest way to perform a photo attack is by holding still the picture in front of the sensor by hand or using a fixed support. The presence of uncanny hand-shaking motions or the complete absence of movements are key discriminant cues that are captured by the proposed method. Experiments on ReplayAttack and MSU databases are conducted to determine if the proposed method is able to separate these unnatural motion cues from natural movements. Three subsets of the ReplayAttack database are considered:

- hand: photo attacks (print and digital) performed by holding the picture by hands.
- fixed: photo attacks (print and digital) performed using a fixed support.
- both: includes hand-held and fixed photo attacks.

For the MSU database, attacks are displayed on a fixed support only but extra camera motion is present when the android sensor is used for authentication. For this reason, acquisitions made with the android sensor and those made with the fixed laptop sensor are considered separately during the evaluation in order to assess the impact of camera motion on the proposed method.

The vocabulary is learned from the whole training set for each database but a specific classifier is trained for each photo attack scenario. Table 3.4 reports the performance of the proposed countermeasure against photo attacks without voluntary movements from ReplayAttack and MSU databases.

Table 3.4: Performance of the proposed countermeasure against photo attacks without voluntary movements.

HTER/EER(%)	ReplayAttack			MSU		
	hand	fixed	both	android	laptop	both
rigid	3.3	1.6	3.1	0	0	2.5
non-rigid	10	4.9	6.8	30	10	35
fusion	1.4	1.0	4.0	15	10	7.5

Analysis of non-rigid motion The absence of facial expression variations is used as a discriminative motion cue between real and fake faces and it is measured by the proposed non-rigid motion features. Experiment on photo attacks from the ReplayAttack database obtains $HTER = 6.8\%$. Errors come from the real accesses that exhibit almost no motion as clients remain still (neutral expression) in front of the camera. Even worse results are achieved on similar photo attacks from the MSU database with $EER = 35\%$ for the same reason. Under such authentication protocols, liveness cues are too subtle compared to the estimation noise of the CLM-parameters and poor results are achieved.

Analysis of rigid motion To cope with limited expression changes during authentication, global head motion is used as a complementary cue to discriminate between natural head movement and fake one. The global face movement can be identified depending on the way the picture is displayed in front of the sensor, ie fixed on a support or holding by hands. In the first case, the picture is fixed and no motion is detected whereas real faces are usually not strictly immobile. Almost perfect detection is obtained against this attack scenario on both ReplayAttack and MSU databases with $HTER = 1.6\%$ and $EER = 2.5\%$ respectively. Looking closer at photo attacks from MSU database, both real accesses and attack attempts recorded by the mobile android sensor exhibit shaking motion due to camera movement whereas samples acquired using the fixed laptop webcam for authentication only present shaking motion on attack attempts. In the second case, hand-held photo attacks are more or less easy to detect depending on the presence of extra hand-shaking motion. Sufficient hand-shaking movements are present on ReplayAttack samples and good performance is achieved with $HTER = 3.3\%$.

Fusion Performance degrades when both rigid and non-rigid motion features are combined compared to the proposed method based on rigid motion only. In the end, rigid motions are more discriminant than non-rigid motion under non-cooperative authentication protocols due to the lack of vitality of certain real accesses as they maintain a neutral face. In this situation, the only remaining discriminant cue corresponds to extra hand-shaking motion measured by rigid motion features.

3.4.1.2 Photo attacks with simulated liveness

To simulate liveness, impostors try new ways to perform photo attacks by simulating natural facial expressions such as eye-blinking and head movements. This type of attack represents a new challenge for motion-based countermeasures and experiments are conducted on CASIA database to cope with this type of attack. Unlike ReplayAttack and MSU databases, CASIA recordings exhibit more freedom during authentication where the user is allowed to smile and talk adding slight expression variations. Impostors either warp/bend/move around printed photos to simulate liveness ('warp') or wear them like a mask to simulate eye-blinks ('eye-cut'). Table 3.5 reports the performance of the proposed countermeasure against photo attacks with simulated liveness from CASIA-FASD.

Table 3.5: Performance of the proposed countermeasure against photo attacks

HTER/EER(%)	CASIA		
	warped	eye-cut	both
rigid	12.3	15.8	15.5
non-rigid	6.7	9.0	8.7
fusion	2.2	6.7	9.0

Non-rigid motion features are able to detect non-rigid deformations of the face region due to unnatural bending movements and yield $EER = 6.7\%$ against warped photo attacks. Also, the absence of mouth movements helps with the detection of photo attacks for this use-case as $EER = 9.0\%$ for eye-cut photo attacks. Failing samples correspond to the situation where almost no movement is present for both attacks and real accesses. When simulated movement is incorporated, rigid movements only achieve 15.5% as no extra hand-shaking movement is noticeable when holding the picture particularly when the impostor is wearing the picture like a mask and rigid motion becomes very similar to real head movements. Nonetheless, unnatural simulated motion is correctly characterized by both rigid and non-rigid motion parameters jointly as increased performance is obtained on warped and eye-cut photo attack detection respectively with $EER = 2.2\%$ and $EER = 6.7\%$. Overall, comparable results are obtained for the method based on non-rigid motion only and the one combining both rigid and non-rigid motion with $EER = 8.7\%$ and $EER = 9.0\%$.

3.4.1.3 Discussion

This evaluation highlights the various nature of discriminant motion cues related to the attack scenario in place and reveals the strengths and limitations of the proposed countermeasure. On the one hand, non-rigid motion features are efficient only for authentication protocols where the client expresses sufficient liveness cues such as eye-blinks and mouth motion. In this configuration, the proposed method is able to detect photo attacks performed with simulated liveness quite well on

CASIA database. On the other hand, rigid motion is discriminant mainly for close-up attacks as shaking motion is amplified as the distance to the sensor decreases. Despite limited movements from certain real accesses, the proposed method based on rigid motion yields almost perfect detection on both ReplayAttack and MSU databases against still photo attacks (no voluntary movements). Besides, the type of sensor (fixed or mobile) used for authentication has almost no impact on the detection and almost perfect detection is obtained on MSU database.

3.4.2 Evaluation of the proposed method against video attacks

To carry out a video attack, the impostor must acquire a video clip of a real client. This can be done directly by installing a hidden camera or by hijacking an online video clip of the targeted client as video sharing services become more and more popular. In practice, the video clip is acquired using a different sensor than the one belonging to the face recognition system. Especially, most of the time video clips are obtained from a hand-held camera whereas authentication is done with a fixed sensor leading to some extra camera motions when performing the video attack. If the impostor uses his/her hands to hold the displaying medium in front of the sensor, additional shaking movement is present. Another motion factor is related to the camera motion introduced during authentication for mobile applications. Consequently, the resulting motion is likely to be different between real accesses and video attack attempts in practice.

3.4.2.1 Experiments on MSU database

Experiments are conducted on video attacks coming from the MSU database. These attacks are displayed on a fixed support but different sensors are used for authentication and for acquiring videos used for spoofing. We consider four different video attack scenarios with an increasing level of difficulty. Fisher encoding is learned from the whole training set whereas a specific classifier is learned to detect video attacks from each of the following scenarios.

- Scenario 1: A fixed camera is used for authentication while spoofing video clips are obtained from a hand-held camera (MSU laptop-iphone).
- Scenario 2: A hand-held camera is used for authentication while spoofing video clips are obtained with a fixed camera (MSU android-ipad).
- Scenario 3: A hand-held camera is used for authentication and spoofing (MSU android-iphone).
- Scenario 4: A fixed camera is used for authentication and spoofing (MSU laptop-ipad).

Table 3.6: Performance of the proposed countermeasure against video attacks from the MSU database.

EER (%)	Scenario 1	Scenario 2	Scenario 3	Scenario 4
rigid	0	0	10	35
non-rigid	10	20	20	10
fusion	10	10	20	10

Performance is reported in table 3.6. Rigid motion features outperform non-rigid features except for scenario 4. Perfect results are achieved on scenarios 1 and 2 as extra camera motion is present in either attacks or real accesses. The proposed method manages to identify rigid movements due to camera shake from rigid movements due to natural head movements. Also, only two miss classified attack attempts are obtained in scenario 3 although extra camera shake is present in both real accesses and attacks. Camera shaking movements are amplified on attack recordings due to the proximity between the sensor and the display medium used for spoofing, which is handled by the proposed method. Rigid movements are unable to separate real accesses from fixed video attacks when both recordings are obtained from a fixed sensor (scenario 4 obtains $EER = 35\%$) as natural movement and replayed natural movements are identical.

Unexpectedly, non-rigid motion features yield decent results on data acquired with the laptop sensor in scenarios 1 and 4 with $EER = 10\%$. One possible explanation is that the difference in face size between real and fake faces impacts the estimation of non-rigid motion parameters. Otherwise, non-rigid motion are only generated by natural expression changes which are also present in video attack attempts and should not constitute reliable motion cues. The fusion of both rigid and non-rigid motion features do not improve the detection overall and only rigid features should be considered for video attack detection.

3.4.2.2 Experiments on CASIA and ReplayAttack databases

To complete our study, additional experiments are conducted on ReplayAttack and CASIA datasets. Performance is reported in table 3.7. Video attack recordings from ReplayAttack DB are similar to scenario 1. Almost perfect detection is achieved ($HTER = 2.5\%$) with rigid motion features. In this experiment, non-rigid motion features yield only $HTER = 18.5\%$ and fusion obtains $HTER = 6\%$. This confirms that non-rigid motions are usually not fitted to detect video attacks.

Experiment on CASIA recordings is similar to scenario 4 but with mid-range video attacks. This attack scenario is the most challenging from a motion perspective because no extra motion is present when recording and displaying the fake face. Poor detection is obtained with $EER = 21.1\%$ for both rigid and non-rigid motion.

Table 3.7: Performance of the proposed countermeasure against video attacks from ReplayAttack and CASIA datasets.

HTER / EER (%)	ReplayAttack	CASIA
rigid	2.5	27.8
non-rigid	18.5	28.9
fusion	6.0	21.1

3.4.2.3 Discussion

In this section, different video attack scenarios are identified along with the associated discriminant motion cues. The difference between the motion of the sensor used for authentication and the motion of the camera used to capture the face for spoofing is crucial for the detection of video attacks. In particular, extra camera motion happen to be very helpful for the detection task. The proposed method based on rigid motion is capable of detecting video attacks almost perfectly when

the sensor used for authentication is fixed whereas the camera used for spoofing is mobile or the other way around. Additionally, decent results are obtained when mobile camera and sensor are used for spoofing and authentication respectively.

The detection of video attacks acquired by a fixed camera and recaptured by a fixed sensor remains a major limitation of the proposed method and motion-based countermeasure in general.

3.4.3 Robustness of the proposed method using different sensors

Experiments on CASIA database are conducted to confirm the robustness of the proposed method to different sensors. IFV are computed on the whole training set but different classifiers are learned to detect photo and video attacks with respect to the low (LD), medium (MD) and high (HD) definition sensors. Performance is reported in table 3.8. Surprisingly, performance varies between the three datasets (one for each sensor) especially for video attacks. After further investigations, it appears that video attacks for both LD and HD datasets are performed using a fixed support to limit the hand motion when holding the iPad. Hand-shake motion itself is too weak to be highly discriminant and only $EER = 26.7\%$ is achieved. Better performance is obtained for the MD dataset with $EER = 10\%$ as no support is used when performing the video attacks. Furthermore, another difference is highlighted by the drop of performance on photo attacks for the HD dataset to $EER = 13.6\%$. This data has been cropped to suppress useless background information and reduce the volume of the data. As a consequence some artificial camera motion is present adding unnatural motion on real accesses. Although this experiment is not able to prove the robustness of the proposed method relatively to the sensor choice, we are able to highlight the extent of the proposed method to detect unnatural motion.

Table 3.8: Impact of the sensor on the proposed countermeasure performance

HTER / EER (%)	photo	video
CASIA (LD)	6.8	26.7
CASIA (MD)	3.3	10
CASIA (HD)	13.6	26.7

3.4.4 Overall evaluation and comparison with state of the art motion-based countermeasures

Additional experiments on the PrintAttack and PhotoAttack database (subsets of ReplayAttack) are carried out for comparison with state of the art motion-based countermeasures. Results are reported in table 3.9. Only two methods outperform the proposed countermeasure on the PhotoAttack database. Perfect detection is achieved by Bharadwaj and al [Bharadwaj13] using HOOF features and $HTER = 1.5\%$ is obtained by Anjos [Anjos14] using face/background motion correlation. However, both methods are computationally expensive due to optical flow computations compared to our method. On the contrary, our method outperforms existing countermeasures on CASIA database with $EER = 18.7\%$. One advantage of the proposed countermeasure over face/background consistency methods is that it can be used for close-up (fake face and background) or mid-range (face only) attacks as background motion is not taken into account. Furthermore, with additional feature normalization schemes (ie: feature normalization between $[-1, 1]$ or PCA),

almost perfect detection can be achieved on ReplayAttack with $HTER = 1.5\%$. In conclusion, the proposed method is highly competitive with state of the art motion-based countermeasures.

Table 3.9: Comparison of motion-based countermeasures on PrintAttack, PhotoAttack, ReplayAttack and CASIA databases. Performance is reported in terms of EER for CASIA database and HTER is used on the other databases.

Reference	Algorithm	PrintAttack	PhotoAttack	ReplayAttack	CASIA	MSU
2011, Lorenzo and al. [Chakka11] (SIANI)	Face parts motion	10.63	-	-	-	-
2011, Anjos and al. [Anjos11, Freitas Pereira13]	Face/background motion correlation	8.98	7.2	11.79	26.65	-
2013, Bharadwaj and al. [Bharadwaj13, Bharadwaj14]	(EVM) + HOOF	**0**	**0**	**0**	21.11	-
2013, Warris and al. [Waris14]	EVM + STACOG	-	-	9.12 [1]	-	-
2014, Anjos and al. [Anjos14]	Face/background motion correlation	-	1.5	-	-	-
	Face parts motion [2]	-	39.4	-	-	-
	Optical Flow Field [3]	-	75	-	-	-
Proposed method	IFV based on rigid motion	7.5	3.3	4.6	20	20
	IFV based on non-rigid motion	7.5	6.8	13.7	**17**	19
	IFV based on rigid and non-rigid motion	3.1	4	5	18.7	**10**

[1] System trained with half the training set
[2] System inspired from [Kollreider07a]
[3] System inspired from [Bao09]

3.4.5 Discussion

Unlike anti-spoofing methods based on face-background motion correlation, the proposed method detects unnatural rigid face motions and extends to video attacks provided that extra shaking motion is introduced by camera motion or by hand-holding the spoofing medium. The proposed method is able to differentiate shaking motion induced by camera motion during authentication and hand-shaking motion introduced when holding the displaying medium. Only rigid motion cues are helpful against video attacks whereas non-rigid motion cues are necessary to detect mid-range photo attacks with simulated motion. For this reason, we proposed a fusion scheme to exploit both type of movements as no prior information on the type of attack is available to select one type of movement over another for testing.

3.5 Conclusion

In this chapter, a complete review of exclusive motion-based countermeasures is provided. A novel motion-based countermeasure is introduced. Rigid and non-rigid motion sequences contain discriminative information for photo and video attack detection. The Fisher framework has demonstrated its ability to build discriminant mid-level features from variable length video sequences as Fisher

vectors are able to characterize unnatural motion by learning an unsupervised codebook of micro movements (short motion sequences).

A detailed analysis of spoofing attacks and their associated discriminant motion cues is given throughout extensive evaluations on the latest public anti-spoofing databases. The way an attack is performed has a significant impact on motion-based countermeasures as extra movements like shaking motion, warping and moving constitute discriminant motion cues. Sufficient vitality signs in real accesses and sufficient hand-shaking motions in attacks are needed for good detection. These conditions are generally satisfied in close-up attack scenarios as demonstrated by experiments on ReplayAttack and MSU databases. Conversely, mid-range attacks are difficult to detect especially when limited movement is available as natural and simulated motion become ambiguous. Certain instances of video attacks are impossible to detect using motion only unless shaking motion is present. Having said that, the proposed method is robust to extra camera motion when using a mobile sensor and is competitive with state of the art methods.

In conclusion, motion-based countermeasures are not able to cope with all attack scenarios but should still be implemented as an additional safety measure in addition to texture based countermeasures to consolidate the fake face detection against replay attacks. The fusion of motion and texture based countermeasures are left to future works.

Chapter 4

Anti-spoofing countermeasures based on the recapturing process model

Contents

In this chapter, the problem of anti-spoofing is addressed by modelling the radiometric and spatial distortions involved in the recapturing process. The primary objective is to understand the transformations between a single face capture and its recaptured version and to look for some consistency from one identity to the other. The second part deals with the application of the measured distortions to discriminate between real and fake faces. Finally, experiments on the synthesis of spoofing attacks for new identities are conducted.

4.1 Related work

Although several works have exploited colour and blur information to detect spoofing attacks [Schwartz11, Tronci11, Galbally14a], Wen et al. are the first that explicitly modelled the distortions involved in the recapturing process as a combination of a low-pass filtering process (blur distortion) and an histogram transformation (colour distortions) of the reflectance component. They proposed three features to account for the blurriness, colour diversity and chromatic distribution between real and fake faces. They also exploited the specular reflections to complete their method. Following their analysis, we demonstrate that the recapturing process can be modelled as a combination of blurring and exposure transformations and we propose a parametric model for both distortions. To estimate the different transformations, enrolment data is used as prior. A similar problem has been addressed by Joshi and al. [Joshi10] who use good quality face exemplars to enhance low quality face pictures using both global image corrections and face specific patch-based corrections. Their global procedure involves deblurring and colour adjustments using a Bayesian framework with prior constraints derived from the good quality exemplars and resemble at what we are trying to do, to model the recapturing process. In their case however, multiple good quality exemplars spanning different pose and illumination conditions are available. In our case, we suppose that the authentication procedure requires a frontal pose with no facial mimics and occurs in a controlled environment with known lightning so only one exemplar is sufficient for recognition and is available for our anti-spoofing countermeasure. Otherwise, only few works involve the use of face recognition enrolment samples for face anti-spoofing. Among them, Chingovska and al. [Chingovska15] propose to build person specific classifiers for every enrolled client to leverage identity specific discriminant information during the classification stage. They prove that common features (LBP, $LBP-TOP$, $motionfeatures$) hold client-specific information even though they are designed to capture spoofing artefacts. Yang and al. [Yang15] also use additional enrolment samples to train person-specific classifiers. Their approach consists in synthesizing virtual features corresponding to the fake face class for each client when no actual attack samples are available. They assume that the relation between two subjects features is a translation plus a linear transformation whereas the relation between real and fake samples for each client is identical. A classifier is then trained for each client using those real and virtual fake face features.

4.2 Capturing versus recapturing processes

In this section we describe the recaptured image formation pipeline and propose a reliable model to measure the radiometric and spatial distortions observed between a real and a fake face. The replay attack face spoofing is illustrated in figure 4.1. First, the impostor needs a first high quality capture of the target real face to manufacture a fake face. Then, this high quality sample is printed or displayed onto the spoofing medium (paper or screen). It is aligned with the camera sensor plane at a certain distance so that the recaptured face maintains a similar size as the real access.

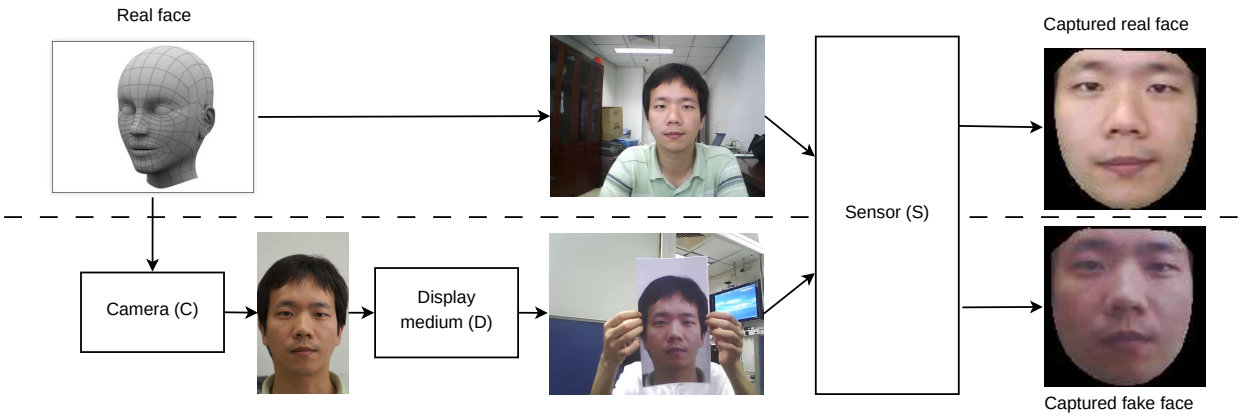


Figure 4.1: Pipeline of the recapturing process

4.2.1 Image capture pipeline

Let us review the main procedures occurring in the image formation pipeline related to optics and exposure camera settings. We break down the analysis from a radiometric standpoint and from an optical standpoint.

4.2.1.1 Radiometry

The complete image sensing pipeline from a radiometric perspective is illustrated in figure 4.2.

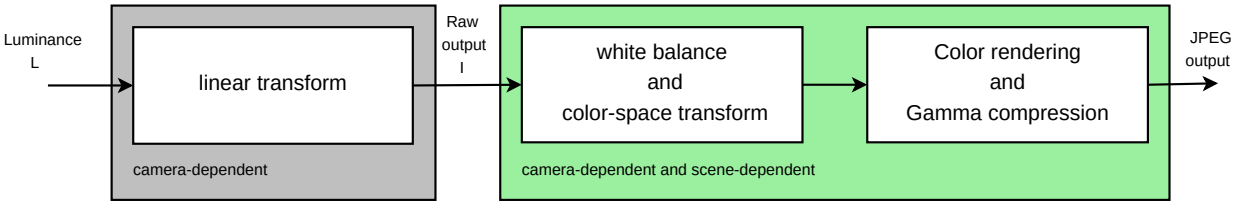


Figure 4.2: Image formation from radiometric perspective.

From a radiometric standpoint, a digital camera RAW output I results from the linear transform of the luminance L of a scene into discrete pixel values depending on the camera settings as described in [Hiscocks11].

$$I_{raw} = \alpha(\frac{tS}{f_s^2})L \tag{4.1}$$

where each term refers to: RAW digital image (I_{raw}), calibration constant for the camera (α), exposure time (t), aperture number (f_s f-stop), ISO Sensitivity (S) and luminance of the scene L. Those parameters control the overall exposure of the scene and are usually set automatically using different camera modes ('portrait', 'landscape','auto'). Then, this raw image follows different digital post processing operations such as demosaicing, sharpening, white balancing and colour rendering before JPEG compression. Following the digital camera model proposed in [Chakrabarti09], pre-processing operations such as flare, noise removal, dark current compensation, quantization, filling marking of dead pixels and compression artefacts are considered as noise and are discarded in first approximation. White balance and internal colour-space transformation used for colour rendering are modelled by a 3*3 orthogonal matrix W. Colour rendering modifies the tristimulus values

(usually in XYZ colour space) so that they fit within the limited gamut and dynamic range of the output colour space (usually sRGB) in a way that is visually pleasant. It differs from one manufacturer to the other and is modelled by a non-linearity transformation $g : R^3 \rightarrow R^3$. The final model is given in equation (4.2):

$$I = g(WI_{raw}) \tag{4.2}$$

White balance and colour rendering are not only camera-dependent but scene-dependent as well. These aspects are exploited to detect the disparities between a real face acquisition and a recaptured face which has been first captured by a different camera under the same illumination conditions.

4.2.1.2 Point Spread Function (PSF)

From an optical standpoint, the image formation maps a real point in 3D space to a 2D pixel. A series of geometric transformations and blurring operations transform a sharp ideal image into a digital image as illustrated in figure 4.3. For this study, we neglect geometric distortions such as chromatic aberrations and vignetting as they appear only on certain occasions and we focus on blur sources only via the point spread function (PSF) model.

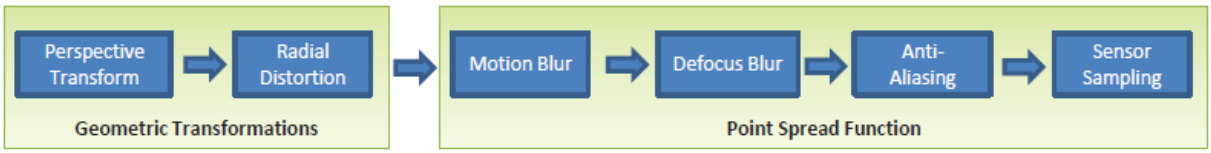


Figure 4.3: Image formation from an optical perspective

Digital cameras capture images by integrating incoming light on their imaging chip. Camera properties such as the fill factor, shutter speed (exposure time), aperture size and lens focal length affect the PSF. The fill factor corresponds to the active sensing area size as a fraction of the theoretically available sensing area. It controls the aliasing effect due to pixel sampling. An anti-aliasing filter (low pass) is usually added to avoid aliasing effects and is responsible for blur observed due to a lack of resolution. Shutter speed and aperture control the amount of light measured by the sensor by adjusting the integration time and aperture size. Both affect the resulting motion blur (long integration time increases motion blur) and depth of field (small aperture leads to a large depth of field). In the end, the resulting blur is a combination of sensor anti-aliasing blur, motion blur and defocus blur.

All blur sources between an ideal high resolution sharp image I_0 and the camera output I is modelled as a global blur kernel K following equation (4.3).

$$I = q(K * I_0) + n \tag{4.3}$$

where I_0 is a super-resolved sharp image, q is a point-sampling operator that matches the size of I_0 with I and n models the noise.

4.2.2 Radiometric distortions involved in the recapturing process

Based on the presented image formation model, we now describe the main procedures involved when capturing a fake face. Some image artefacts such as aliasing, banding effects, vignetting and any compression artefacts are not considered at first approximation because they are inconsistent between attack scenarios.

4.2.2.1 Luminance produced by fake faces

In addition to exposure adjustments made by each capturing device (high definition camera and authentication sensor), the colorimetry of the recaptured face is determined by the optical properties of the medium used for spoofing. Each medium has a different display mechanism to convert a digital image back to luminance.

- **Digital display** Digital screens are calibrated so that colours are represented accurately on the monitor. The luminance output L_s of a given digital image I is determined by three major settings namely 'brightness control (b)', 'contrast control (c)' and 'gamma control (γ)' following equation (4.4):

$$L_s = cI^{\gamma} + b + rA \tag{4.4}$$

  where rA corresponds to the reflected ambient light on the screen. b has a minimum value corresponding to residual light still emitted from the screen (0 pixel value) as individual liquid crystals cannot completely block all light from passing through. Maximum value is determined by the power of the backlight system. LCD screens act as direct light sources compared to reflected light from real scenes. When performing an attack, contrast and brightness values must be adjusted so that no on screen reflections are visible that is $cI^{\gamma} + b >> rA$

- **Print copy** Perceived luminance L_s of a printed digital image I is the result of ambient light A (uniform on the paper surface) reflected upon the paper surface with reflectance $R = r * I + r_0$ where r and r_0 embed the reflective properties of the paper. $r_0 * A$ accounts for the minimum reflected light corresponding to pure black ink in a given ambient illumination A. Higher image contrast can be obtained by using good quality paper such as copper/mate paper. Equation (4.5) gives the observed luminance:

$$L_s = (rI + r_0)A \tag{4.5}$$

  The same model can be applied to masks as they follow the same light diffusion mechanisms but with different optical properties.

The above equations are given for gray scaled images and can be extended to colour images as both equations apply on each colour channel independently. Combining (4.4) and (4.5), a general formulation of the conversion of pixel values in luminance values for each colour channel by both photo paper and digital displays is given by:

$$L_{s|r,g,b} = s_{|r,g,b} I^{\gamma_{|r,g,b}} + b_{|r,g,b} \tag{4.6}$$

where $s_{|r,g,b}$ is a colour scaling factor, $b_{|r,g,b}$ accounts for the colour bias (minimum brightness) and $\gamma_{|r,g,b}$ encodes the non-linearity ($\gamma_{|r,g,b} = 1$ for print attacks) on colour channel R,G or B. To simplify the notations, we drop the colour index in the rest of the manuscript.

4.2.2.2 Recaptured luminance

The face luminance L is first captured by a high definition camera giving a digital image. Then this sample is printed or displayed on screen and acquired by the authentication sensor leading to the recaptured image I_f as illustrated in figure 4.1. The resulting transfer function between L and I_f is given by:

$$I_f = g_s(W_s.s.(g_c(W_c L))^{\gamma} + b) \tag{4.7}$$

where s denotes the colour scaling (3x1 vector), b corresponds to the colour bias (3*1 vector) and γ models gamma correction (3*1 vector) from the spoofing medium. The non-linearity transforms from the authentication sensor (S) and the high definition camera (C) are modelled by g_s and g_c respectively. W_s and W_c account for the white balance and colour-space transform by both sensors.

To identify the radiometric transformations between a real face sample I_r and its recaptured version I_f, we first simplify the camera model by approximating the general linear colour transform W to be diagonal and the non linearity as a simple gamma function on all three channels $g(x_r, x_g, x_b) = [x_r^{\gamma_r}, x_g^{\gamma_g}, x_b^{\gamma_b}]$. After some basic derivations, equation 4.6 can be rewritten to link I_r and I_f in a per-channel form as:

$$I_{f_i} = (c_i I_{r_i}^{\gamma_i} + b_i) \tag{4.8}$$

In this simplified modelling, we have demonstrated that the colour distribution of real accesses and recaptured faces varies in terms of contrast, brightness and white balance. To detect if a face is real or fake, we propose to estimate those radiometric distortions in section 4.3.

4.2.3 Blur involved in the recapturing process

During the spoofing process, a high definition face sample is either printed on a given paper or displayed on a given screen. In both cases, cropping or re-sampling the image is required to match the limited paper size or screen size. Hence, the spatial definition of the displayed face corresponds to the definition of the screen (ppi) or the printer (dpi). This down-sampling can be modelled as a combination of low pass filtering and point-sampling. At last, we approximate the recapturing process as a combination of the different blur sources as follows:

$$I_f = q(K_{s2} * G * K_{s1}) * I) + n \approx q(K_{s2} * G) * I) + n \tag{4.9}$$

where K_{s1} and K_{s2} are the PSFs of the high definition camera and the authentication sensor respectively, G is the Gaussian blur due to down-sampling, n models the noise, I is a super-resolved ideal acquisition of the face and q is the point sampling operator which matches the size of I with I_f's. In practice, limited motion is observed during the authentication as the person holds still in front of the camera so the high definition camera blur is totally masked by the lower definition display medium. Similarly, when the resolution of the authentication sensor is higher than those of the spoofing medium, the resulting blur mainly comes from the screen anti-aliasing and the motion of the medium during the acquisition. Real accesses and recaptured samples exhibit different motion blurs particularly in close-up attack scenarios as shaking motion and defocus are magnified as we get closer to the sensor. In order to determine if a sample is a real access or an attack, we propose to recover its blur kernel and to extract its magnitude and shape as it contains information about the type of movement and the eventual low pass filtering due to the limited definition of printers or screens.

4.2.4 Summary

First, we proved that radiometric distortions exist between real faces and fake ones even though the face data used for manufacturing the fake face is captured under the same illumination conditions as real authentication attempts.

Second, we have analysed the different blur sources present in replay attacks. Blur results from a combination of defocus, motion blur and screen down-sampling. When manufacturing the attack,

the impostor is able to obtain a high quality face picture of a valid user to manufacture a fake face. Hence, the blur source comes from the screen or printer anti-aliasing filter to fit the digital image on the physical support, from motion and from defocus during authentication.

4.3 Application to anti-spoofing

In the previous section, the recapturing process involved when performing a replay attack has been studied and discriminative cues have been highlighted. We now present how to estimate these distortions and how they can be applied for anti-spoofing purposes. The problem of anti-spoofing comes down to recovering the blur kernel from equation (4.9) and the radiometric transform from equation (4.7). In practice, both equations are solved using heuristic methods because the true luminance of the scene and the ideal sharp image are unknown. In this work, we take advantage of prior information on each client thanks to available enrolment samples. The proposed anti-spoofing countermeasure is placed after the face recognition system and is designed to detect attacks once the identity of the client is checked. We suppose that the matching works perfectly so that for each client we can pair up the actual face acquisition with the enrolled face sample contained in the face recognition database. Under these assumptions, for each authentication attempt a pair of images is formed using the observed image and the enrolment image corresponding to this client. The radiometric and spatial transformations between both images are computed and used as features for detection. We focus on the face region exclusively mainly to maintain background independence as much as possible. Faces are registered using eyes locations to compensate for head rotations and cropped using a bounding box determined by the interocular distance. The general pipeline is presented in figure 4.4.

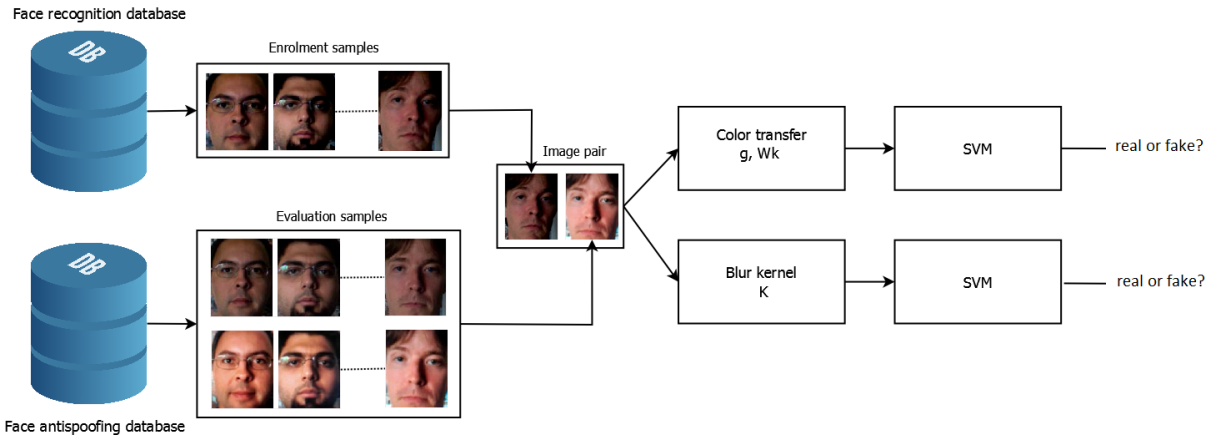


Figure 4.4: Pipeline of the proposed method

4.3.1 Radiometric distortions estimation

The radiometric transformations between a fake face and its real counterpart involve a combination of colour scaling, colour offset and gamma non linearity in the case of spoofing attacks as described by equation (4.8). This equation relies on simplifying assumptions and more complex transformations are likely to be missed. Hence, we consider two parametric models for the radiometric transformations between an authentication attempt and its corresponding enrolment sample with an increasing level of complexity:

- **per-channel Gamma model**: the radiometric distortion f is modelled by a per-channel gamma transform followed by an affine transform. This model yields 3 parameters per colour channel for a total of 9 coefficients.

$$f_i(x) = \alpha_i x^{\gamma_i} + \beta_i \ , \ \ i = [r, g, b] \tag{4.10}$$

- **coupled Gamma model**: the radiometric distortion function f is modelled by a per-channel gamma transform followed by a general affine transform. This model yields a total of 15 coefficients. This model allows the coupling of colour channels and is better suited to account for colour-space and white-balance transformations.

$$f([x_r, x_g, x_b]) = [r^{\gamma_r}, g^{\gamma_g}, b^{\gamma_b}] * C + [\beta_r, \beta_g, \beta_b] \tag{4.11}$$

  The first model is an easy one where each coefficient relates to color transformations identified in the recapturing process. The second model is more general.

The estimation of the model parameters is done in three steps. First, the global colour transfer approach proposed in [Pitie05] is employed to produce an auxiliary image A that has the texture of the reference image I_{ref} with the colour distribution of the observed sample I_{test}. To estimate the color distortions between the two, only low frequencies are necessary so we extract the base layer of both images. Finally, the radiometric transformation f is estimated using non-linear least square regression.

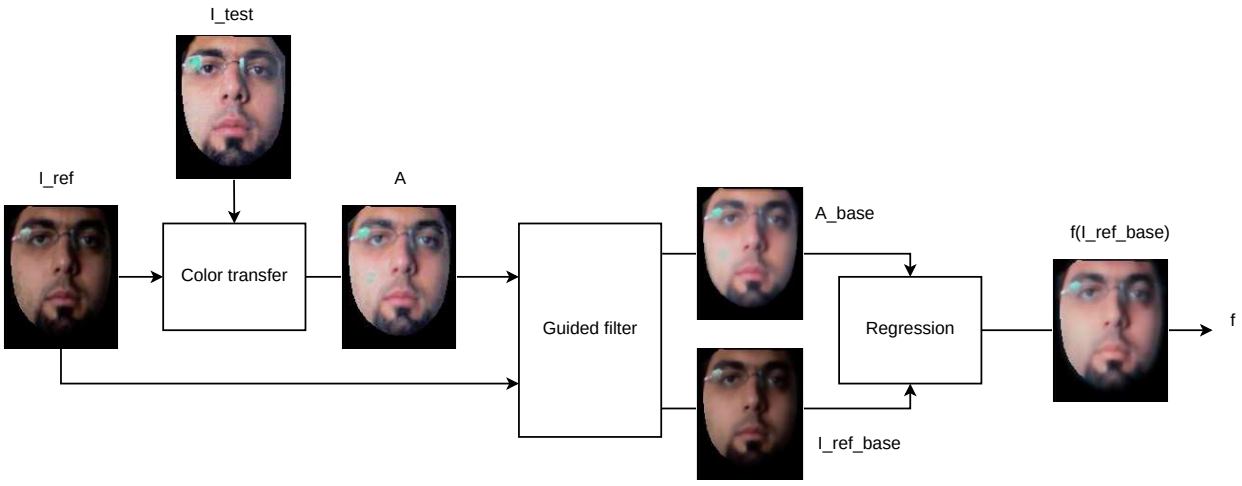


Figure 4.5: Pipeline of the proposed method to recover radiometric distortions between a test sample I_{test} and its reference I_{ref}.

4.3.1.1 Colour transfer

The colour transfer method of Pitie et al [Pitie05] consists in transforming the original colour pdf into the target colour pdf by breaking down the 3-dimensional problem into a succession of 1-dimensional pdf transfers for which a simple solution is available (see equation (4.12)).

$$t(x) = C_Y^{-1}(C_X(x)) \tag{4.12}$$

where C_X and C_Y denote the cumulative histograms of grayscale image X and grayscale image Y respectively. Random projections are used to improve the pdf transfer iteratively by analogy with Radon transform. After 30 iterations, the algorithm converges and we obtain an auxiliary image that has the same texture as the enrolment sample but with the colours of the observed sample.

4.3.1.2 Illumination component extraction

By nature, color and contrast information corresponds to low frequency image components. We adopt a texture/base layer image decomposition to separate image details from the illumination component. Only the base layer is retained to estimate the color transformations between the original reference image I_{ref} and the color-transferred image A following equation:

$$A_{base} = f(I_{ref_{base}}) \tag{4.13}$$

This way, undesirable effects generated by the colour transfer procedure (high frequency artefacts) are avoided and a reliable estimation of the colour transformations is achieved. The base layer extraction is performed using a low pass edge preserving filter. The Guided Filter (GF) [He13] is employed in this work to extract the base layer of both I_{ref} and A by using I_{ref} as the guidance image to preserve the original image texture. Two critical parameters control the filter design, the radius of the filter neighbourhood r and the smoothing index ϵ. The patches with variance much smaller than ϵ are smoothed, whereas those with variance much larger than ϵ are preserved. These two parameters are selected experimentally, the values of $r = 2$ and $\epsilon = 0.01$ provide good results and are kept for all the experiments.

4.3.1.3 Non-linear regression

The second step then consists in the estimation of f's parameters using non-linear regression techniques. The gamma model is solved using an iterative approach by minimizing the fitting error in the mean square sense using the "lsqcurvefit" function from Matlab. This function uses the trust-region reflective algorithm with full finite differencing to approximate the gradient of the objective function so we don't need to explicitly provide the Jacobian.

4.3.1.4 Classification

Those nine parameters (three for each colour channel) are used for fake/non fake classification. The coefficient of determination r^2 associated with the regression on each colour channel is also added in the feature set to account for the goodness of fit of the proposed model. Indeed, many complex recapturing mechanisms have been omitted especially specular reflections which generates some fitting errors especially in high intensity regions. We recall the definition of the r^2 coefficient in least square regression. Let $(y_i)_{i=1..n}$ represent the set of image values for data $(x_i)_{i=1..n}$ and $(f_i)_{i=1..n}$ the estimated values of y_i using least square regression. The R^2 coefficient can be seen as the fraction of unexplained variance defined by:

$$r^2 = 1 - \frac{\sum_{i=1}^{n}(f_i - y_i)^2}{\sum_{i=1}^{n}(y_i - \bar{y})^2} \quad \text{with } \bar{y} = \frac{1}{n}\sum_{i=1}^{n} y_i \tag{4.14}$$

In the end, 12 parameters constitute the proposed colour features, 9 model parameters and 3 r^2 coefficient on each channel. We also define the R^2 coefficient as the average of r^2 on all three color channels.

4.3.2 Blur estimation

Recovering the blur kernel from equation (4.9) is a classic problem in blind deconvolution. A sharp estimate of the observed (blurred) sample is required to recover the blur kernel. In our case, enrolment samples are used to approximate the sharp version of the observed samples. Inspired by the method of [Pan14] for blind deconvolution with exemplars, we recover the blur kernel K in an iterative minimization procedure. Their idea is to use exemplars to estimate salient edges $\nabla_s E$ accurately from which a first estimation of the PSF is achieved by solving (4.15)(replacing ∇S by $\nabla_s E$). This estimation is then refined using alternate minimization between a latent sharp image S (4.16) and the blur kernel K in the gradient domain (4.15). A conjugate gradient method is used to solve (4.15) while a half quadratic splitting L0 minimization method is required for (4.16).

$$\min_k \|\nabla S * K - \nabla I\|_2^2 + \lambda_1 \|K\|_2^2 \tag{4.15}$$

$$\min_S \|S * K - \nabla I\|_2^2 + \lambda_2 \|\nabla S\|_0 \tag{4.16}$$

where λ_1 and λ_2 are two regularization constants. We set the blur kernel size at 15*15 pixels. To have a compact set of features, the blur kernel is approximated by a 2D Gaussian function. The amplitude, the standard deviation along x, the standard deviation along y and the orientation of the Gaussian constitute the proposed 4-length blur feature vector. These blur features are fed to an SVM classifier for classification.

4.4 Evaluation of the radiometric distortion model

First, the validity of the proposed method for the estimation of the radiometric distortions is verified. Then, anti-spoofing experiments are conducted on ReplayAttack, CASIA and MSU databases.

4.4.1 Model validation

The estimation of the radiometric distortions between a real face and its fake counterpart is a three step process involving color transfer, low pass filtering and regression. Experiments on the Replay-Attack, CASIA and MSU databases are conducted to validate each process and to select between the per-channel Gamma model and the coupled Gamma model. The coefficient of determination R^2 is used to measure the goodness of fit for both models (here the R^2 corresponds to the average of r^2 on each channel to have a single coefficient per video). We display the distribution of R^2 for the ReplayAttack, CASIA and MSU databases in figure 4.6. Both models are able to characterize the radiometric distortions generated by the recapturing process as the average R^2 is close to one for all three databases. We also observe that higher R^2 is achieved using the coupled Gamma model as it has more parameters and thus better fit the data.

To illustrate that the proposed method provides a good estimation of the radiometric distortions between real and fake faces, samples with the lowest R^2 coefficient are shown in figure 4.7 in columns (d) and (e). The first two columns represent the corresponding real and fake faces respectively. The column (c) displays intermediate results corresponding to the output of the color transfer between the real face and its fake counterpart. First of all, these exemplars provide qualitative proof of the discriminative power of the radiometric distortions generated by the recapturing process. Secondly, these distortions are transferred efficiently to the real face image using the color transfer procedure of Pitie and al. [Pitie05]. Looking at CASIA exemplars in detail, specular reflection artefacts

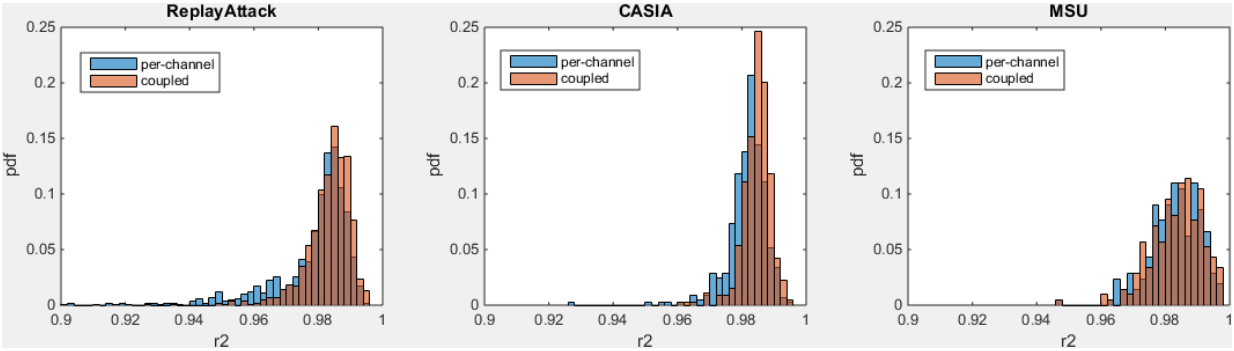


Figure 4.6: Distribution of the R^2 coefficient for the per-channel Gamma model and the coupled Gamma model on the ReplayAttack, CASIA and MSU databases.

present during the attack are also transferred to the new coloured real face image. The next step estimates the parameters of the distortion model using least square regression between the original real face image (a) and the new coloured image (c). In the second and third rows, we observe that regression errors can result from unexpected image artefacts due to the recapturing process such as light reflections (second row) or color fading effects (third row). For this reason, we also consider the R^2 coefficient as an extra feature. A slight color bias is observed on the first and third rows between the color transfer output (c) and the regression result using the per-channel Gamma model (d) showing the limitations of the per-channel model. On the contrary, the coupled Gamma model is able to fit almost perfectly the radiometric distortions.

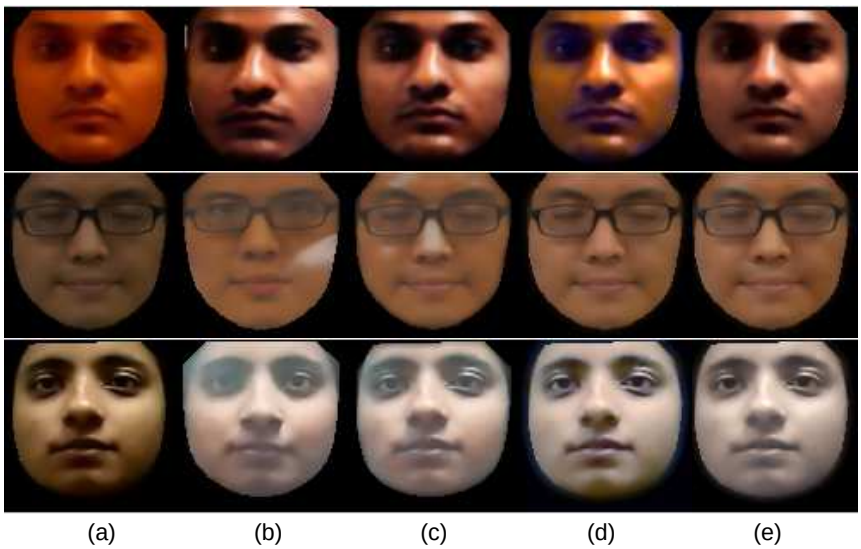


Figure 4.7: Regression results corresponding to attacks with the lowest regression coefficient (R^2) from ReplayAttack (first row), CASIA (second row) and MSU (last row) databases. From left to right: (a) enrolled samples, (b) attack samples, (c) color transfer outputs, (d) regression results using the per-channel Gamma model, (e) regression results using the coupled Gamma model.

4.4.2 Analysis of model parameters

The complex radiometric transformations between real and fake faces are approximated by a combination of colour scaling α, colour bias β and γ non-linearity on each colour channel when considering the per-channel Gamma model. This simple model allows a good understanding of the recapturing

process and is used to describe the exposure and color modifications specific to print attacks and projected attacks (using screen). This model is used to investigate the consistency of radiometric transformations for each type of attacks and for multiple identities. For a given database, the probability density functions (PDFs) of each model parameter is derived per attack type. Results on ReplayAttack, CASIA and MSU databases are shown in figures 4.8, 4.9 and 4.10 respectively. Three general observations are drawn out from these plots:

- The offset parameters $\beta = [\beta_r, \beta_g, \beta_b]$ are inferior to $1/255 = 0.039$ except for very few samples so it is neglected and set to zero.
- The consistency of radiometric transformations between real and fake faces depends on the attack type. Colour changes generated by print recaptures are more consistent then fake faces displayed on screens as the PDFs of the model parameters follow narrower distributions. Especially, the non-linearity parameters γ follow a Gaussian centred between 0.5 and 1 which confirms the low contrast impression when looking at print attack samples. Radiometric changes for digital attacks (using screens) induce color scaling α to be greater than one as they often appear brighter (screens are direct light sources).
- The radiometric transformations involved in the recapturing process are not consistent between identities especially for digital attacks as indicated by the large PDFs of the model parameters. It appears that face skin color has a significant impact on the radiometric transformations due to the recapture. This is in line with the image formation pipeline presented in section 4.2.1 as scene-dependent transformations such as white balance occurs. Nonetheless, these parameters can be used as classification features as they are highly discriminant overall.

Looking at parameters from the coupled Gamma model, we notice that the C matrix can vary significantly between similar samples. From a classification perspective, this variability is a serious problem as we want features that are consistent between two similar samples. After further investigation, it appears that the high correlation between the three channels leads to an ill-posed estimation problem. Further constraints are required to regularize the solution, one way is to force some orthogonality constraints on C and add a diagonal matrix D that embeds color scaling on each channel independently. For these reasons, we consider only the per-channel Gamma model for the rest of the experiments.

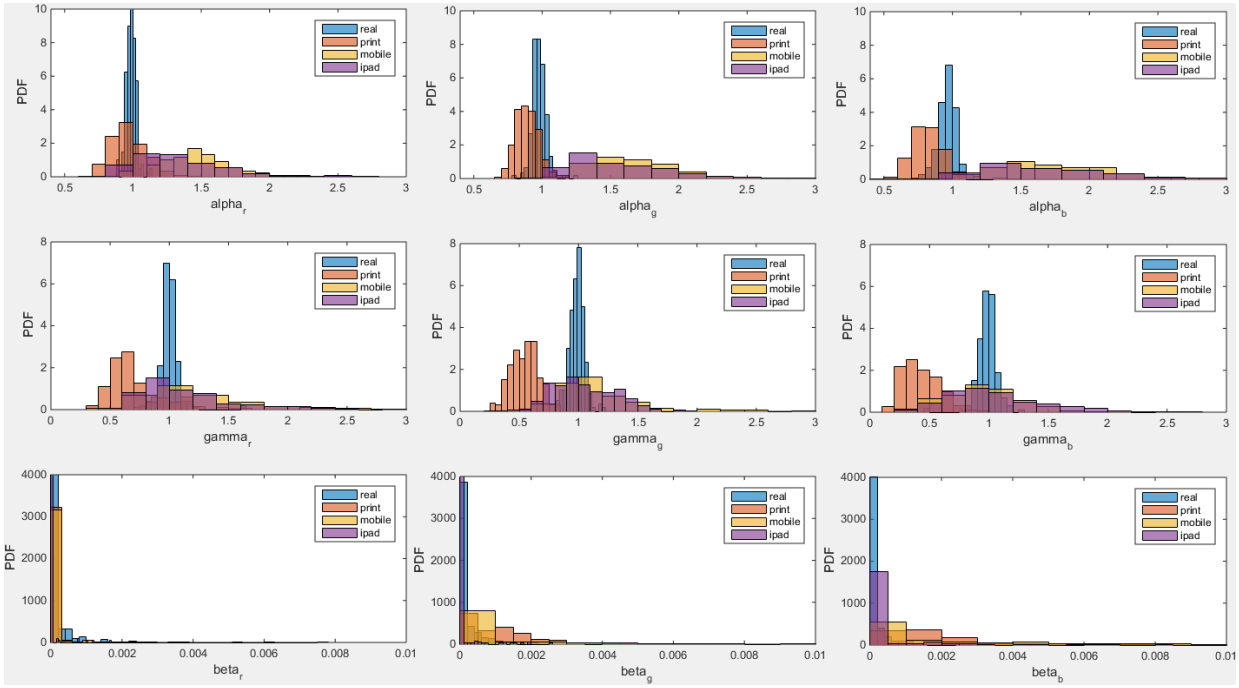


Figure 4.8: PDFs of the per-channel Gamma model parameters across the ReplayAttack database for real faces, print attacks, mobile attacks and iPad attacks.

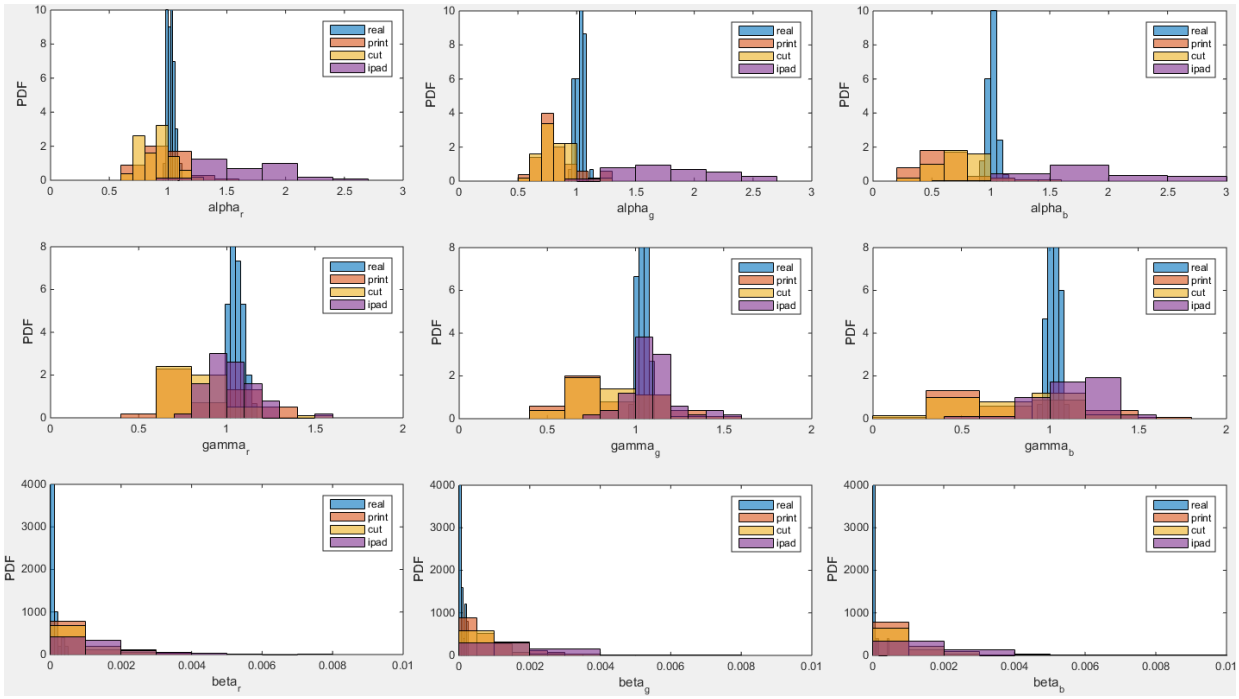


Figure 4.9: PDFs of the per-channel Gamma model parameters across the CASIA database for real faces, print attacks, cut attacks and iPad attacks.

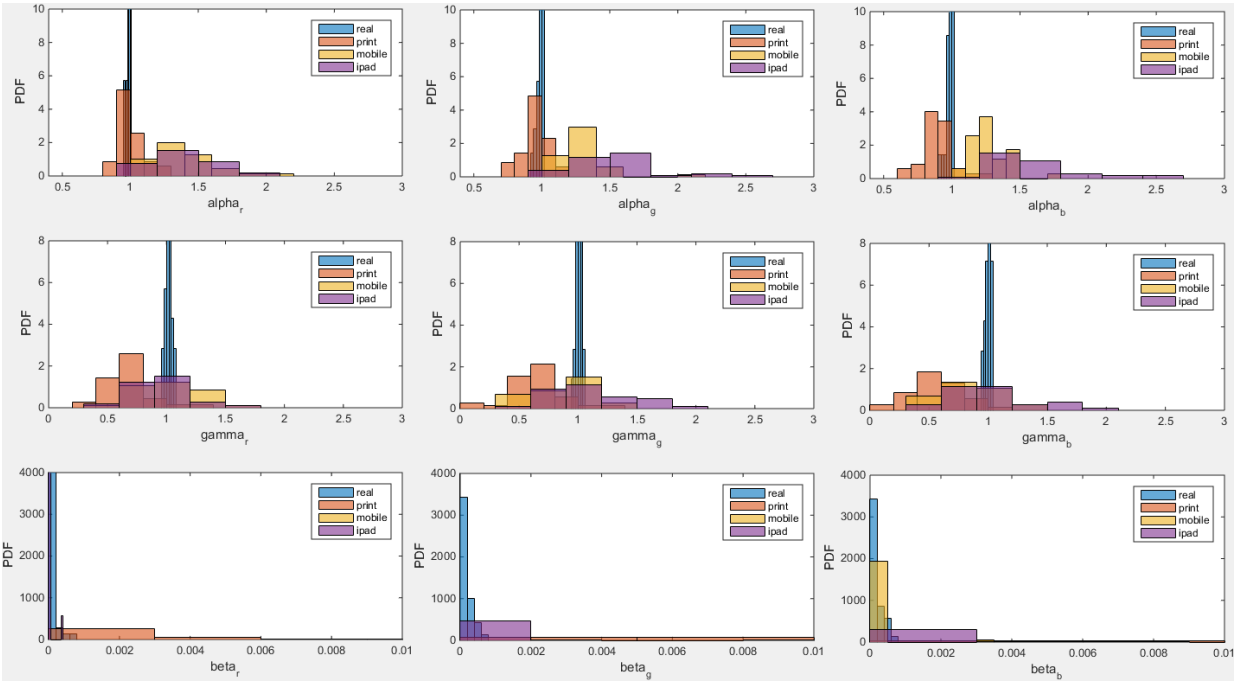


Figure 4.10: PDFs of the per-channel Gamma model parameters across the MSU database for real faces, print attacks, mobile attacks and iPad attacks.

4.4.3 Classification results

Four different experiments are conducted to study four use-cases that are more or less restrictive in terms of implementation.

Experiment A The first experiment investigates the case where acquisition conditions are fixed so enrolment and authentication are performed under the same conditions. In that case, color variations between the authentication attempt and the corresponding enrolled sample directly reflects the radiometric distortions due to the recapturing process. Evaluations on the ReplayAttack, MSU and CASIA (HD) datasets are performed and classification results based on the proposed colour features are reported in table 4.1.

Table 4.1: Detection results of experiment A.

Experiment A	MSU (EER)		CASIA (EER)	Replay-Attack (HTER)
	android	laptop	HD camera	LD camera
print	0	0	3.3	1.9
mobile	0	0	-	0.3
iPad	0	0	6.7	1.9
overall	0	0	3.3	0.7

Almost perfect results are obtained on ReplayAttack and CASIA databases with $HTER = 0.7\%$ and $EER = 3.3\%$ respectively while perfect results are obtained on MSU database. Looking at the errors on CASIA, only one real access is completely miss classified and should be considered as an outlier. Errors on ReplayAttack are caused by one iPad photo attack and one print attack

that have similar exposure as their real counterparts. Because enrolment data for MSU and CASIA are made from the same video as real accesses, the results are overly optimistic. Nonetheless, this experiment shows that radiometric distortions are highly discriminant regardless of the quality of the sensor and regardless the type and quality of attacks. However, this setting is very restrictive and is hardly applicable in real world applications because illumination are likely to change from one authentication attempt to another.

Experiment B The second experiment supposes that faces are enrolled under several illumination conditions that corresponds to typical authentication use-cases. For a given authentication attempt, the system first retrieves the enrolled sample from the claimed identity and this sample matches the actual illumination conditions of the authentication attempt. Although the system has not been trained under these new illumination conditions because manufacturing fake faces under multiple illuminations is costly, the proposed method can cope with new lightning settings provided that the retrieved enrolled sample presents similar illumination as the actual authentication attempt. For this purpose, the ReplayAttack database is selected for the evaluation. Evaluation is conducted by training the system on samples acquired under the 'controlled' lightning and by testing on samples obtained in 'adverse' lightning. The pairing of the authentication attempt and the corresponding enrolled sample satisfies the illumination correspondence assumption. The proposed method yields perfect detection in this testing configuration. This indicates that the relative color distortions between a real and fake face are consistent between two illumination conditions. This result is overly optimistic because in practice there are slight variations in illumination between enrolment conditions and authentication ones.

Experiment C The third experiment investigates how performance degrades when there is a mismatch between illumination of enrolment samples and the actual illumination conditions. We consider the low quality and high quality datasets from the CASIA database for this experiment. Enrolment samples and fake face samples are acquired under controlled illumination conditions using the real accesses acquired from the high quality sensor. However, authentication is performed under uncontrolled illumination using samples captured by the low quality sensor and the anti-spoofing system is trained and tested using this data. Using the ideal case where enrolment samples match the illumination of authentication attempts (experiment A), we obtain $EER = 3.3\%$. When enrolment and authentication have different illumination conditions, the EER drops to 20%. In this case, the classifier misinterprets color variations due to the recapturing process and color variations due to illumination changes. Looking at the resubstitution results (training with all the data), the EER reaches 2%. This suggests that both types of color variations can be discriminated using the proposed features.

Experiment D Generalization to new types of attacks is a major quality for anti-spoofing countermeasures. Indeed, spoofing attacks evolve step by step and new ways to break through any security system are to be found sooner or later. Thus, the capability of the proposed countermeasure to detect attacks that have not been considered during the training process is investigated. To evaluate the robustness to future attacks, cross-protocol evaluations are performed on ReplayAttack and MSU databases. We investigate three scenarios where the system is trained using two out of the three available types of attacks, while the third one serves for testing. The scenarios descriptions are as follows:

- Scenario 1: train with mobile and iPad attacks, test on printed photographs.

- Scenario 2: train with printed photographs and iPad attacks, test on mobile attacks.
- Scenario 3: train with printed and mobile attacks, test on iPad attacks.

Table 4.2 shows the efficiency of the proposed method to generalize on unseen attacks. Mobile and iPad attacks are well detected even without training the classifiers with either one. This comes from the fact that both mobile and iPad attacks are similar by nature (attack on screen) so training with either one is good enough. For the print attack scenario (1), the detection error increases significantly because features are very different between print and digital attacks as the recapturing mechanisms are also different as explained in section 4.2. Nonetheless, the proposed method achieves encouraging results.

Table 4.2: Generalization of the proposed method to unseen types of attacks

colour features	MSU (EER)		Replay-Attack (HTER)
	android	laptop	LD camera
scenario 1	15	5	8.1
scenario 2	0	0	3.1
scenario 3	0	0	5.9

4.4.4 Overview

First, we have demonstrated that the recapturing process produces exposure and color changes compared to a single image capture even when spoofing face data is obtained in the exact same illumination conditions as authentication. These radiometric transformations are well modelled by the proposed per-channel and coupled Gamma models which are good approximations of the radiometric transformations generated by the recapturing process. The per-channel Gamma model is retained for its parameter estimation consistency although the coupled-model offers slightly better fitting properties. It is generic enough to embed different color transformations related to the variety of spoofing attacks. Also, the r^2 coefficient is used as an extra feature as it reflects fitting errors due to specular reflections or other non-modelled artefacts.

Second, the implementation the proposed method in a real application is only limited by the amount of enrolment data. A dense set of enrolment samples that spans the set of illumination conditions encountered during authentication is required. The system is able to discriminate between a real face and a fake one only if the associated enrolled sample and real access are close enough which should be the case if the enrolment set is dense enough. In this case, the proposed method provides a good metric to assess if the authentication attempt comes from a real access or from an attack by modelling the radiometric differences in terms of scaling, color offset and non-linearity.

4.5 Evaluation of the PSF model

In addition to radiometric distortions, the recapturing process also generates blur distortions as mentioned in section 4.2.3. In this section, we study the different blur sources found in Replay-Attack, CASIA and MSU databases for different sensors and different types of spoofing attacks. First, the estimation of the PSF is discussed. Then, classification results are presented.

4.5.1 Model validation

In equation 4.9, we have identified the different blur sources involved in the recapturing process as a combination of three PSFs (the authentication sensor, the sensor used for spoofing and some Gaussian blur generated when fitting the digital image on the displaying support). For both sensors, the PSF results from defocus, motion blur and sensor anti-aliasing (limited resolution). In practice, the impostor can easily obtain a high quality picture of the face to spoof. The quality of the reproduction is then limited by the displaying medium size. For example, when displaying a 12 Mp image (4000*3000) on an iPhone, the screen automatically down-samples the digital image to the screen resolution (480*320). Similar down-sampling occurs when printing a picture as printer and paper quality have limitations in the number of pixels it can reproduce per inch (ppi). For a given paper size, the professional standard uses a printing setting of 300 ppi which forces the resolution of digital images to a given resolution to fit in the paper so the printer software re-samples digital images to this size before printing. Table 4.3 shows examples of required image resolution for printing on standard paper format. The authentication sensor needs sufficient resolution power to detect this additional blur source. Defocus blur can also reinforce the overall blur as close-up attacks may get out of focus. Also, motion blur adds up and may mask out the other discriminant blur sources if motion is too important. Nonetheless, we believe that motion blur can also be discriminant as typical fake face motion such as shaking or warping may generate different motion blurs from real face motions.

Table 4.3: Examples of image resolutions for printing at 300 ppi.

Paper format	required image size
A3	3456*5184 (18Mp)
A4	2400*3300 (8Mp)
A5	1740*2460 (4.2Mp)

Note: pixel per inch is often confused with dot-per-inch (dpi) to characterize the printing resolution. The dpi denomination is a term used by manufacturers to advertise their product as more dpi is assimilated as a quality factor although it depends on the printing technology (two printers with the same printing resolution (ppi) can have different dpi in function of the technology).

Synthetic blur To investigate if the PSF is correctly recovered by the method of [Pan14], experiments are conducted on synthetic blurred image samples. We select one typical image from the CASIA database and apply a series of Gaussian blurs with increasing size going from $\sigma = 0.5, 1, 2$. To recover the PSF, salient edges are manually obtained from the original image which is used as the reference image for the de-blurring process. The PSF size, sparsity prior weight λ and the number of iterations are set empirically to 15, 0.002 and 10 respectively. We approximate the recovered PSF by a 2D general Gaussian (asymmetric) to assess the similarity between the original Gaussian blur and the recovered PSFs in terms of blur strength. A toy example with a more complex blur simulating shaking motion is also presented. PSFs results are illustrated in false color in figure 4.11.

The retrieved PSF approximately captures the general form of the synthetic blur kernel but strong horizontal and vertical artefacts are present leading to a slight over estimation of the blur strength. These examples reveal some severe limitations of the method for a reliable estimation of complex blurs. Nonetheless, this method is used in anti-spoofing detection for exploratory purposes

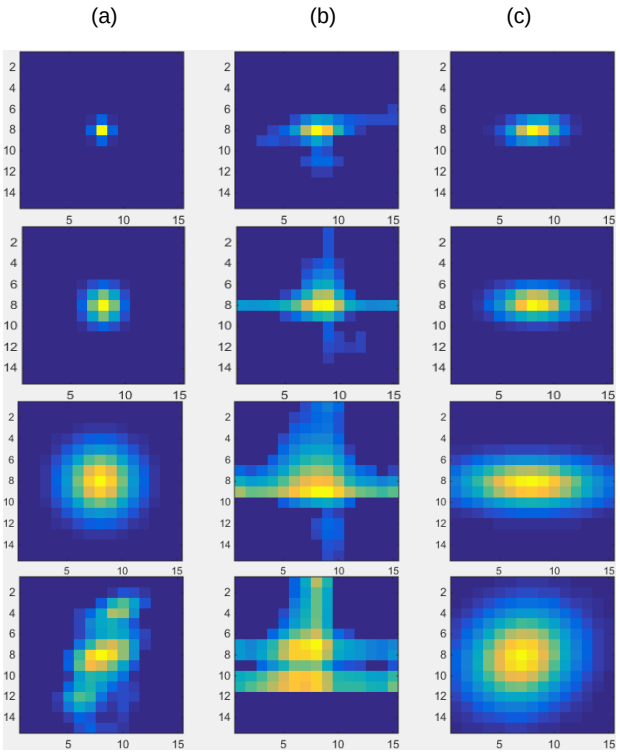


Figure 4.11: Blur estimation on synthetic images. From left to right: column (a) depicts the original blur source, column (b) shows the recovered PSFs and column (c) displays the Gaussian approximation of (b). From top to bottom: Gaussian blur with $\sigma = 0.5$, Gaussian blur with $\sigma = 1$, Gaussian blur with $\sigma = 2$ and complex blur.

as the method is still able to recover the blur strength. Features are extracted from the Gaussian approximation of the kernel. The amplitude, the blur strength (σ) in both principal directions and the residual error R^2 of the Gaussian fitting procedure are retained.

4.5.2 Classification results

Quantitative evaluations are performed on Replay-Attack, MSU and CASIA databases to assess the discriminative power of the blur features. Table 4.4 reports anti-spoofing performance of the proposed method using the proposed blur features.

Table 4.4: Performance of blur features

colour features	MSU (EER)		CASIA (EER)	Replay-Attack (HTER)
	android	laptop	HD camera	LD camera
print	25	40	36.7	14.4
mobile	15	30	-	15.3
iPad	35	10	6.7	34.7
overall	30	40	30	44.1

The best performance achieves $EER = 6.7\%$ and is obtained for video attacks using a high definition camera from the CASIA database. For this scenario, the gap between the resolution

of the screen and the resolution of the authentication sensor is large enough to reveal the screen down-sampling blurring effects. Interestingly, the proposed method is not working well on printed attacks. Even though print attacks have few details, the remaining ones are sharp and the recovered PSF is similar to those of real faces. An example is given in figure 4.12.

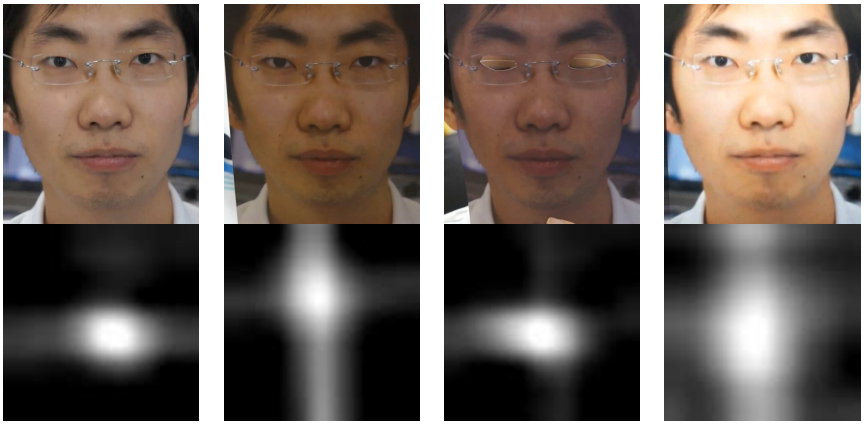


Figure 4.12: Blur kernels estimated by the proposed method on CASIA examples. From left to right, real access, print attack , print eye-cut attack and iPad attack are displayed.

Decent results are obtained on the ReplayAttack database for both printed ($HTER = 14.4\%$) and mobile attacks ($HTER = 15.3\%$). For this database, attacks fake the full scene (face and background) and are performed closer to the sensor. The low quality of the sensor masks out the down-sampling blur. However mobile attacks are performed really close to the sensor due to the small screen size so that defocus blur is observed and captured by the proposed method. On the contrary, the recovered PSFs of printed attacks are narrower than the PSF of real accesses in average. The proximity of the fake face to the sensor when performing a full view attack leads to a better fake face resolution compared to real accesses. In addition, printed faces exhibit texture details due to printing artefacts (banding effects) which makes them look sharper than real accesses. An illustration is given in figure 4.13.

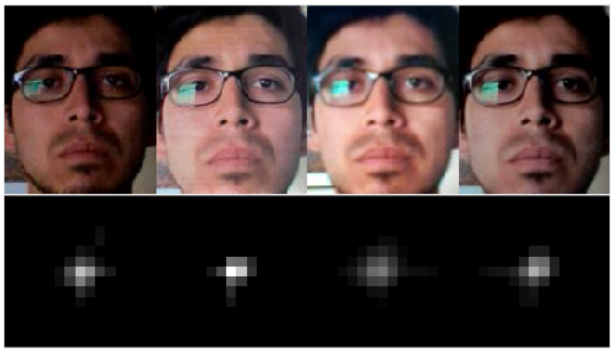


Figure 4.13: Blur kernels estimated by the proposed method on ReplayAttack examples. From left to right, real access, printed attack, mobile attack and iPad attack are displayed.

Results from the MSU database are inconsistent and harder to interpret as both sensors have totally different results. After looking at the image samples, it appears that there are a lot of scaling variations during the authentication process as shown in figure 4.14. As blur is scale sensitive, inconsistent blur kernels are obtained across each authentication attempt.

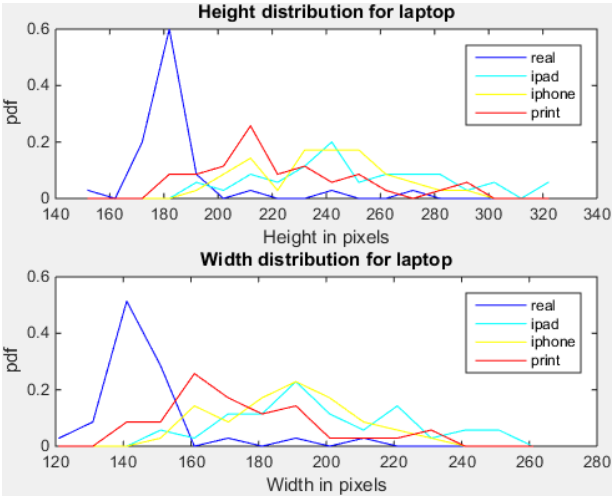


Figure 4.14: Face size distribution for the laptop acquisitions of MSU database

4.5.3 Overview

Even though the proposed blur kernel estimation method only provides a rough estimation of the PSF, general observations can be drawn out. Motion blur is very subtle for both real accesses and replay attacks as video acquisitions are taken (short exposure time and motion compensation) and limited motion is allowed during authentication. Defocus blur mainly occurs when replay attacks are performed really close to the sensor to fake the whole scene using small screens (smartphone screens) but can easily be circumvented using larger screens. At proper distance, actual sensors manage a good focus and the only blur present is the one generated by screens when the resolution of the authentication sensor is superior to the screen resolution. Print attacks are usually sharp and the lack of details is perceived as a texture difference rather than a blurring effect.

Finally, the proposed blur features are efficient under three requirements. First, face acquisitions must have a fixed size for each authentication attempt which requires the user to stand at a fixed distance from the sensor. Second, the sensor needs a sufficient resolution to resolve the blur generated by the screen down-sampling as motion blur and defocus blur are very limited in practice. Third, enrolment samples must have equal or superior resolution compared to the authentication sensor.

A better blur estimation technique is needed to solve for an accurate PSF and exploit subtle motion blur. Even though results are far worse compared to the proposed colour features, blur contains discriminant information to detect digital replay attacks provided that a high quality sensor is employed.

4.6 Synthesis of spoofing attacks

One of the major problem in anti-spoofing is inter-person variability. Despite extensive efforts to find features that are stable from one identity to the other, popular features in anti-spoofing contain client specific information. To take advantage of this aspect, authors in [Chingovska15] propose to learn a client specific classifier using attacks from clients in the training set for new clients (in test set). Authors in [Yang15] go one step further and propose to directly synthesize LBP features for

unseen clients (out of the training set) from enrolment data. Inspired by these works, we look into a new way to synthesize spoofing attacks from enrolment samples. Motivated by the success of the color transfer procedure [Pitie05] in transferring radiometric distortions to enrolled faces, we try to forge artificial fake faces in terms of radiometry. Hence, we focus on the synthesis of the base layer component of face images. Further work is required for the synthesis of the texture layer of spoofing samples and is not treated in this work.

4.6.1 Motivations

Radiometric transformations are client specific distortions as skin color has a significant impact on white balance and image corrections involved in the recapture process. From one identity to the other, the distortions model varies significantly for a given sensor and a given attack type as shown in the analysis of model parameters in section 4.4.2. Our goal is to predict these transformations for new identities based on a limited training set. Let f represent the mapping between a real face and its fake counterpart for a given attack type. This mapping is highly non-linear across different identities with different skin color but if two identities have similar color distributions we expect to have similar distortion transformations. Hence, our idea is to find a local coordinate system such that any new identity can be expressed as a linear combination of similar identities whose spoofing attacks are known. These known identities can be seen as anchor points and the spoofing attack of the new identity is approximated by the linear combination of the spoofing attacks of anchor points. For this, we make use of the color transfer procedure to build a dictionary of anchor points that has the color distributions of known identities for real and spoofing attacks with the spatial texture of the new identity. Then, sparse coding with positivity constraints is employed to find the local representation.

4.6.2 Base layer synthesis pipeline

For a given type of attack and acquisition conditions, the goal is to predict the fake face associated to an enrolled client based on a limited set of training identities (only the base layer is considered in the context of this thesis). Let $X = [x_1, .., x_n]$ and $Y = [y_1, .., y_n]$ represent the set of vectorized face images (base layer) of real accesses and attacks associated to the n identities of the training set. Let x be the enrolment sample corresponding to a new identity. The color transfer procedure of [Pitie05] allows us to transfer the color distribution of training identities to x forming a set of new images $D_x = [d_{x1}, .., d_{xn}]$ and $D_y = [d_{y1}, .., d_{yn}]$. Then, sparse codes α are computed to reconstruct x using the dictionary D_x following the minimization problem:

$$\begin{aligned} \underset{\alpha}{\text{minimize}} \quad & ||x - D_x\alpha||_2^2 \\ \text{subject to} \quad & 0 \leq ||\alpha||_1 \leq \lambda \end{aligned} \tag{4.17}$$

Positivity constraint is added to enforce positive contributions from anchor points. This avoid situations where compensation phenomenon occurs and some anchor points contributions cancel each other. This problem is solved using a modified version of the LARS algorithm to handle the positive constraint using the SPAMS[1] toolbox [Mairal09]. The synthesis of the spoofing attack y corresponding to the client in x is straightforward using the reconstruction with D_y:

$$y = D_y\alpha \tag{4.18}$$

[1] `http://spams-devel.gforge.inria.fr/`

The full pipeline is illustrated in figure 4.15. Face images are normalized to 100*100 pixels for fast computation as we are only interested in the base layer of face images and the regularization parameter λ is set to 0.1.

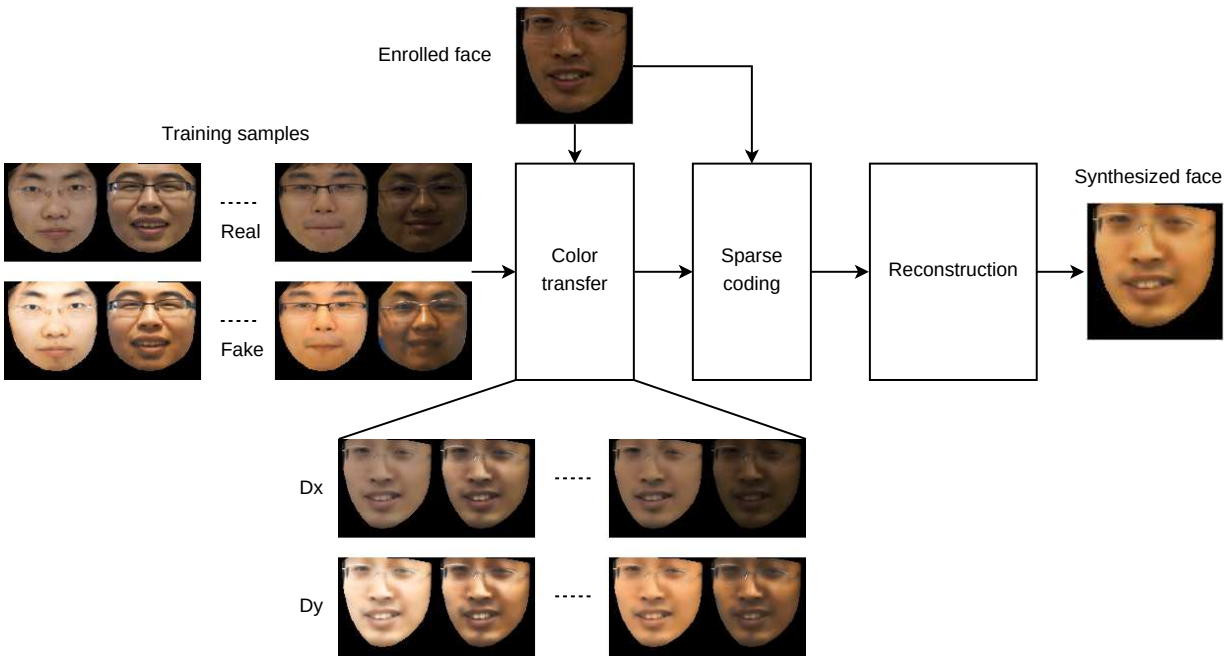


Figure 4.15: Pipeline of the synthesis method

4.6.3 Qualitative Evaluations on CASIA database

Qualitative evaluations are conducted on high resolution acquisitions from the CASIA database only as the same illumination conditions are satisfied between each acquisition only for this subset. In MSU database, background varies between acquisitions of different clients which affects the overall color balance of face samples in an uncontrolled manner. In ReplayAttack, unusual orange color shifts are observed on a few samples acquired under controlled conditions and some fake faces are not detected by the OpenFace toolkit as mentioned in Chapter 2. Consequently, we decided not to experiment on both ReplayAttack and MSU databases.

The proposed method supposes that we can find samples with similar color distributions from the training set. For this reason, we augment the size of the training set compared to the public protocol to make sure that our approximation is valid by selecting 40 individuals out of the 50 available. Images resulting from the synthesis of print attacks (base layer only) are illustrated in the last row of figure 4.16. The first two rows correspond to enrolled samples from clients in the test set and their associated print fake face (ground truth) respectively. The proposed color synthesis method obtains a good approximation of the illumination component (base layer) of spoofing attacks except for a few examples in columns 4 and 9. Besides, specular reflections are not synthesized as seen from exemplars in columns 1 and 3. These specular reflections are generated when photos are bent creating highlights at certain viewing angles and are not consistent from one photo attack to the other and the impostor can avoid them easily. Similar observations are made on the synthesis of eye-cut photo attacks and video attacks in figures 4.17 and 4.18. It appears that radiometric distortions involved when performing video attacks are more consistent from one identity to the other and nice results are achieved.

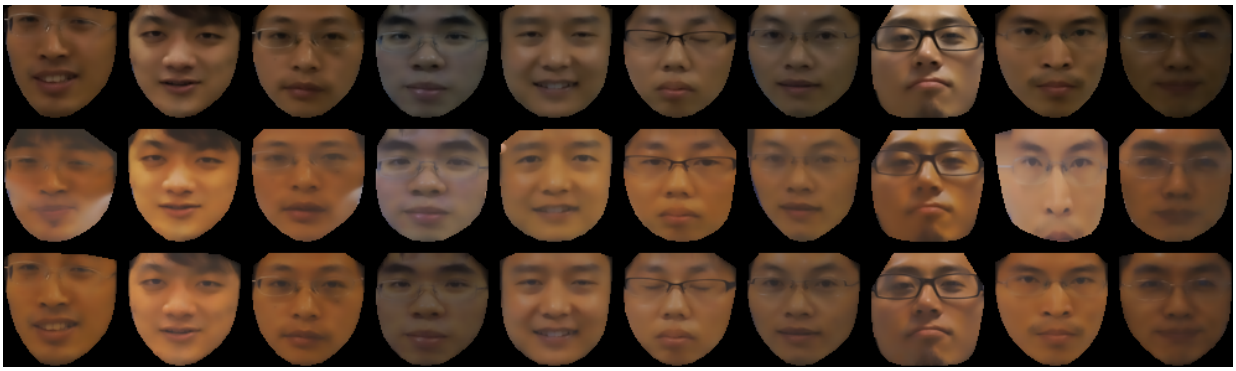


Figure 4.16: Results of the synthesis of print attacks corresponding to high quality acquisitions from the CASIA database. First row corresponds to enrolled faces from the testing set. Second row corresponds to spoofing attacks ground truth. Third row displays synthesized fake faces using the proposed method.

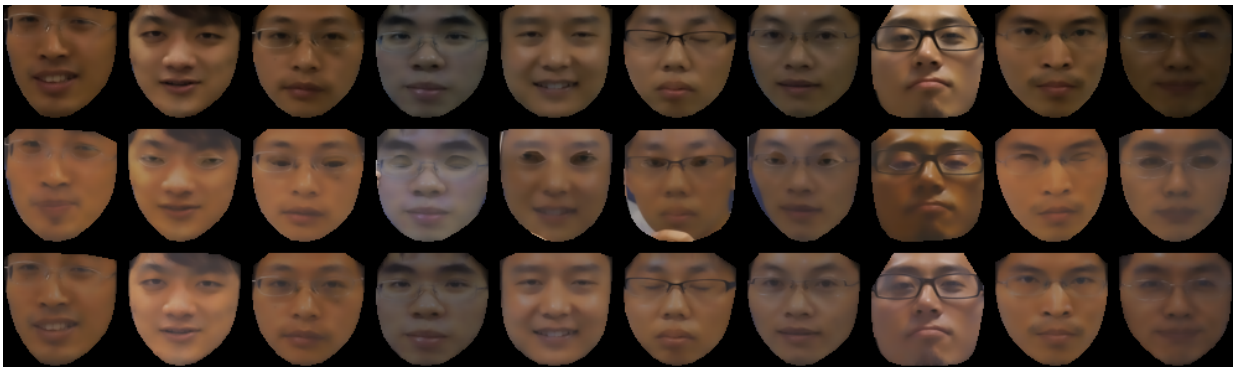


Figure 4.17: Results of the synthesis of eye-cut printed attacks corresponding to high quality acquisitions from the CASIA database. First row corresponds to enrolled faces from the testing set. Second row corresponds to spoofing attacks ground truth. Third row displays synthesized fake faces using the proposed method.

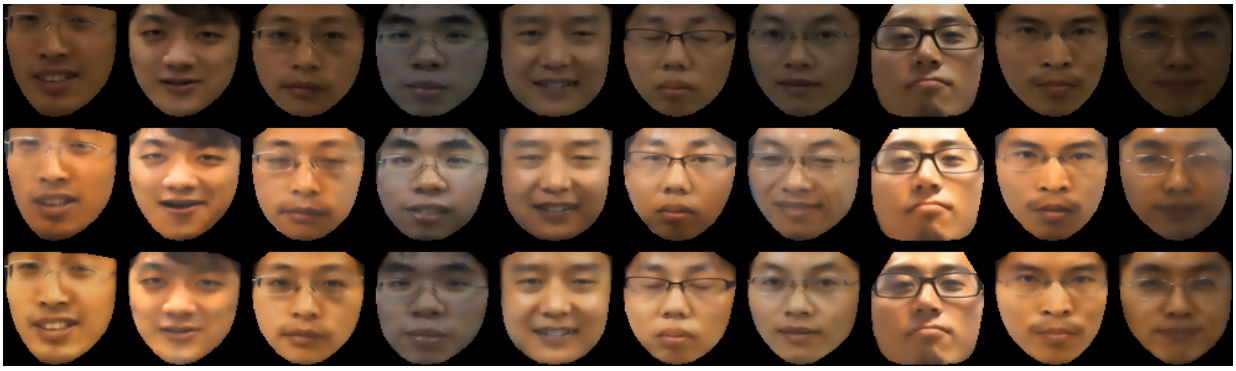


Figure 4.18: Results of the synthesis of iPad video attacks corresponding to high quality acquisitions from the CASIA database. First row corresponds to enrolled faces from the testing set. Second row corresponds to spoofing attacks ground truth. Third row displays synthesized fake faces using the proposed method.

4.6.4 Limitations

Additional experiments have been conducted on low quality samples from the CASIA database without much success because illumination conditions vary randomly between two real access samples on the one hand, and even for a given identity between real accesses and attacks on the other hand. This inconsistency in illumination conditions disrupt the consistency of radiometric distortions for identities with similar skin color. In addition, as mid-range attacks are performed, the changing background also affects the overall image color balance. As a consequence, inconsistent radiometric transformations occur from one identity to the other. Figure 4.19 illustrates this problem on the synthesis of print attacks for the low quality acquisitions of CASIA database. This highlights that the synthesis of spoofing attacks for new identities for a given sensor requires fixed illumination conditions.

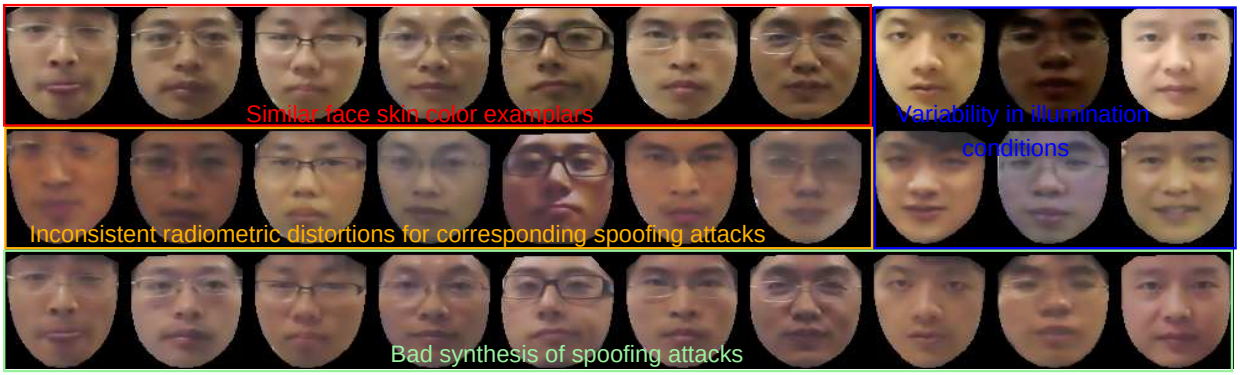


Figure 4.19: Results of the synthesis of print attacks corresponding to low quality acquisitions from the CASIA database. First row corresponds to enrolled face from the testing set. Second row corresponds to spoofing attacks ground truth. Third row displays synthesized fake faces using the proposed method.

4.6.5 Discussion

One of the difficulty of in face anti-spoofing is the high variability between different identities. Although state of the art methods strive for client independent discriminant features such as texture, quality loss, blur or motion information, client-specific information is usually retained and generalization to new identities is difficult. For this reason, a per-client classification approach is advertised and spoofing attacks are required for all clients of the face recognition system. In this section, we have investigated if color distortions generated by the recapturing process can be predicted for new identities regardless of their skin color. For a given sensor and under fixed illumination conditions, we have presented a method capable of synthesizing radiometric distortions on new identities based on color transfer and sparse coding. Although additional experiments on datasets including a wider variety of ethnic groups are needed, the encouraging results on high quality acquisitions from the CASIA database prove the potential of the proposed approach. The key requirement is that some consistency between acquisitions is respected. Not only illumination conditions must stay the same but a complete control over the sensor settings is necessary to have consistent image formation mechanisms especially color balance. Having access to raw images can alleviate some context variability due to rendering operations embedded in the image formation pipeline.

The next step is to synthesize high frequency details of fake faces. Our first trials based on

semi-coupled dictionary learning were unsuccessful and further development are required to make it work. In addition, this study has leveraged two other interesting research directions. First, for a given identity is it possible to predict spoofing attacks under new illumination conditions? Second, is it possible to adapt spoofing attacks to different sensors? These are open questions which should be addressed in future works.

4.7 Conclusions

A detailed analysis of the recapturing process is presented in order to better understand differences between real and fake faces. In first approximation, we demonstrate that radiometric and blur distortions are induced by the recapturing process and we propose a parametric model for both distortions. A novel approach based on the use of enrolment samples to estimate blur and radiometric distortions is proposed. Recovered parameters are used as discriminant features for classification.

- A compact set of 16 features that directly embeds the radiometric differences between real and fake faces is derived. Obtained features relate to physical mechanisms involved in the recapturing process which makes them highly discriminant. The proposed method reliably recovers the radiometric differences between real and fake faces and achieves almost perfect detection on the ReplayAttack, CASIA and MSU databases. It is robust to unseen attacks and is able to cope with slight illumination variations between enrolment and testing.
- The estimation of the blur distortion is not precise enough to exploit subtle blurs. Nonetheless, the proposed blur features obtain decent results on mobile attacks from the ReplayAttack database and on iPad attacks from CASIA high quality dataset as the anti-aliasing blur generated by the displays are resolved by the respective sensors of both databases. Two main requirements are highlighted:
  - the gap between the face resolution from the sensor pixel grid and the face resolution from the screen pixel grid should be wide enough.
  - faces are acquired at the same viewing distance for each authentication to limit face size variability.

In addition, the success of the radiometric distortion model leds us toward the synthesis of spoofing attacks. We proposed a method for predicting the radiometric distortions for new identities under fixed acquisition conditions. Encouraging results are obtained in this direction but further investigations are necessary to completely synthesize a fake face in terms of texture details.

Chapter 5

Certification of face biometric systems

Contents

Biometric systems today are widely used in areas that require a certain level of security and assurance about the used technology. Classical examples for such applications include access control systems to high security areas (like power plants or data centers) and border control systems. Those areas usually require a high degree of assurance in that the used technology is operating as specified and as needed to obtain a secure system. In order to achieve this assurance, independent evaluations and certifications are carried out for the important components of a system or the whole system. The de facto standard for evaluations and certification of components and systems in the area of Information Security are the Common Criteria for Information Security evaluation. A methodology has been developed for the certification of fingerprint biometric systems in 2007. Our contribution within the BIOFENCE project is to apply this methodology to 2D face recognition systems and to assess if this methodology reflects correctly their resistance against spoofing attacks. Eventually, new propositions to adapt this methodology to grade protected 2D face biometric systems with anti-spoofing countermeasures against spoofing attacks.

First, the existing certification methodology designed for the evaluation of fingerprint based systems is presented. Then, we describe how this methodology applies to face anti-spoofing systems. Finally, a practical use-case scenario is investigated where protected and unprotected 2D face recognition systems are rated following the Common Criteria methodology.

5.1 Description of the certification methodology

This section describes the rating approach for the resistance of biometrics systems. This approach is based on the Vulnerabilities Analysis according to Common Criteria Methodology, defined in the 3.0 version of supporting document guidance of Fingerprint Evaluation Mechanism [CCN11]. The evaluation of the vulnerabilities of a biometric system is conducted in two phases:

- Identification phase: corresponds to the effort required to create the attack, and to demonstrate that it can be successfully applied to the biometric system (including setting up or

building any necessary test equipment). The demonstration that the attack can be successfully applied needs to consider any difficulties in expanding a result shown in laboratory to create a useful attack. One of the outputs from Identification could be a script that gives a step-by-step description of how to carry out the attack – this script is assumed to be used in the exploitation part.

- Exploitation phase: corresponds to achieving the attack on a given face biometric system in its exploitation environment using the techniques defined in the identification part. The technique (and relevant background information) could be available for the exploitation in the form of a script or set of instructions and could be performed by a different attacker than the one in the Identification phase. This type of script is assumed to identify the necessary equipment and, for example, mathematical techniques used in the analysis.

In each phase, Common Criteria are rated to quantify the vulnerability of biometric systems against attacks in terms of time, expertise, general knowledge on the system, equipment and accessibility. The ratings of each factor is summarized into table 5.1.

1. **Elapsed Time:** In the Identification phase, it corresponds to the time required to create the attack, and to demonstrate that it can be successfully applied to biometrics system. Applied to spoofing attacks, elapsed time in identification corresponds to the time spent to find the so called "golden fake" and to define a way to build it. For example, for a spoofing attack on fingerprints, it corresponds to the time required to create a spoof from an image of a print (and not the acquisition of this image which is taken into account in the 'Access to biometrics characteristics' factor) obtained with or without the collaboration of the user. "Golden fake" is defined as the best spoof having the best chances to be accepted by the system. In the exploitation phase, Elapsed Time corresponds to the time necessary to apply the "script" to a specific biometrics. For example, the number of trials required to spoof the system contributes to the time rating. Potential difficulties to have an access to the face biometric system in exploitation environment are taken into account in the 'Window of opportunity' factor.

2. **Expertise:** This factor refers to the level of proficiency required by the attacker. A suggested rating for this metric is:

   - Layman: no real expertise needed, any person with a regular level of education is capable of performing the attack.
   - Proficient: some advanced knowledge in certain specific topics (biometrics) is required as well as good knowledge of the state-of-the-art of attacks. The person is capable of adapting known attack methods to his/her needs.
   - Expert: a specific preparation in multiple areas such as pattern recognition, computer vision or optimization is needed in order to carry out the attack. The person is capable of generating his/her own new attacking algorithms.
   - Multiple experts: the attack needs the collaboration of several people with high level expertise in different fields (e.g., electronics, cryptanalysis, physics, etc.). It has to be noticed that a specific competence in biometrics is not considered as "multiple expertise".

3. **Knowledge of the face biometric system:** This factor refers to the amount of knowledge required about the system to perform the attack. For instance, format of the acquired samples, size and resolution of acquisition systems, specific format of templates, but also specifications

and implementation of countermeasures are knowledge that could be required to set up an attack. This information could be publicly available at the website of the sensor manufacturer or protected (distributed to stakeholders under NDA or even classified inside the company). Ratings are:

- Public: information is fairly easy to obtain (e.g., on the web).
- Restricted: information is only shared by the developer and organizations which are using the system, usually under a non-disclosure agreement.
- Confidential: information is only available within the organization that develops the system and is in no case shared outside it.
- Critical: information is only available to certain people or groups within the organization which develops the system.

Note: Special attention should be paid in this point to possible countermeasures that may be implemented in the system and whether it is necessary or not to have knowledge of their existence in order to be successful in a given attack.

4. **Window of opportunity:** This factor refers to the level of accessibility to the biometric system. In identification, it assesses the difficulty to access the system and experiment attack trials in order to find a successful breach in the system. For instance, restricted distribution of the face biometric system to institutions (no distribution to individuals) complicates the design of successful attacks. In exploitation, it reflects the degree of freedom for performing the attack by taking into account:

   - the authentication protocol: multiple trials may be required in a challenge-response authentication protocol.
   - environment: operating conditions related to specific use-cases influence the difficulty to perform an attack such as the presence of a surveillance system, a third party user, and so on.
   - system architecture: the design of the system can facilitate or not physical access to the system for hacking.

   The associated ratings are:

   - Unlimited access
   - Easy
   - Moderate
   - Difficult

5. **Equipment:** This factor quantifies the quality of the equipment required to perform an attack. The corresponding ratings are:

   - None
   - Standard
   - Specialized
   - Bespoke
   - Multiple Bespoke

6. **Access to biometrics characteristics:** This factor corresponds to the difficulty to obtain the biometric information required to make a fake face. For instance, in the case of fingerprint, the fingerprint of a valid user can be recovered directly on the sensor after his authentication attempt using an adhesive film or plastic bag full of water on the scanner. Another possible scenario considers the complicity of the client which helps with the generation of a gummy finger. Because this criteria is very difficult to evaluate objectively as it depends on many external parameters, this criterion is out of the scope of this document and is not taken into account during the application of the certification methodology and we suppose that an impostor is able to obtain the biometric information necessary for the manufacturing of fake face.

Table 5.1: Ratings of the security level of biometric systems

Criteria	Ratings	Identification	Exploitation
Elapsed Time	$\leq$ one day	0	2
	$\leq$ one week	1	4
	$\leq$ one month	3	6
	$\leq$ 3 months	5	8
	$\leq$ six months	7	10
	$>$ six months	10	∞
Expertise	Layman	0	0
	Proficient	2	2
	Expert	5	4
	Multiples experts	7	6
Knowledge of the system	Public	0	0
	Restricted	2	1
	Sensitive	4	3
	Critical	6	5
Window of Opportunity	Unlimited access	0	0
	Easy	1	4
	Moderate	3	6
	Difficult	5	8
	None	∞	∞
Equipment	None	0	0
	Standard	1	2
	Specialized	3	4
	Bespoke	5	6
	Multiple Bespoke	7	8

In order to restrict the number of attacking possibilities which largely depends on a great amount of external factors that may influence the success chances of a given attack, all the ratings and descriptions given in this document are made under the assumption of the worst case scenario. For instance, in the case of attacks with gummy fingers we will consider the existence of a "golden fake", manufactured with a specific material, which, once identified (in the identification phase), is able to break a given scanner/system with very few attempts for almost all the cases. Following the same principle, we will consider fingerprints as public data which can be obtained in a fairly easy manner. This way the rating will start when the attacker has already acquired (by some means) the fingerprint of the user.

The case of attacks involving direct threats to the legitimate user of a given system (e.g., access gained at gunpoint), or violent acts (e.g., attacking a fingerprint verification system with a dismembered finger), falls out of the scope of this document as these actions do not reflect the

security level of a given technology, but rather depend on the willpower of the attacker and are not considered by the CC norm.

5.2 Application to 2D face recognition systems

The main goal is to determine if the certification methodology firstly defined for fingerprint based recognition systems can be used for face recognition system certification. The direct application of the certification methodology is presented on two representative use-cases as examples. The first use-case evaluates the resistance of a non-protected face verification system against video attacks whereas the second use-case considers a system with motion and texture-based countermeasures.

5.2.1 Certification criteria adaptation for face based system assessment

Table 5.2 details the adaptation of each criterion in the identification phase. In this case, the evaluation is impacted by the type of attack and the considered face verification "use-case". For the exploitation phase, the certification methodology is simplified because ratings are identical for all types of spoofing attacks as no particular skill is required to present the fake face correctly with respect to the authentication protocol. Table 5.3 presents the adapted criteria for the exploitation phase. The ratings associated to the Common Criteria during the exploitation phase tend to low security values against face spoofing attacks regardless of the level of protection of 2D face recognition systems. Only the window of opportunity criteria contributes to the security level in the exploitation phase but it depends mainly on the context of use of the face biometric system. In the end, the final security rating of 2D face recognition systems comes from the identification phase and depends on the context of use, the type of attack and the protection measures.

Table 5.2: Identification phase: adapted Common Criteria.

Criteria	Description
Elapsed time	• Time to manufacture fake faces (we suppose that face data from valid clients are already retrieved) • Time to identify a successful attack sample (golden fake face) from experiments
Expertise	• Expertise to manufacture the fake face • Expertise to find the golden fake face
Knowledge of the system	• Information on the sensor (traditional camera, infra-red camera, ...) • Information on the authentication protocol • Information on countermeasures
Window of opportunity	Level of accessibility to the system (is it a commercial product easy to obtain?)
Equipment	Equipment for manufacturing a fake face

Table 5.3: Exploitation phase: adapted Common Criteria.

Criteria	Description	Rating
Elapsed time	Time required to present the fake face correctly, following the authentication protocol	Presenting the fake face is immediate so the rating is "less than 1 day"
Expertise	Level of expertise to use the fake face	Some acting job may be required to mimic real face motion in case of motion-countermeasures but no special skills are required so the rating is "Layman".
Knowledge of the system	Knowledge on the system for authentication	Only the authentication protocol is required to use the fake face so the rating is "Public".
Window of opportunity	Level of access to present the fake face	This criteria depends heavily on the use-case and the rating varies from "unlimited access" for face verification on laptop to "difficult" in an airport security check point.
Equipment	Equipment for performing the attack	No additional equipment except the fake face is needed so the rating is "None"

5.2.2 Application to protected and unprotected face recognition systems

The second goal of that study is to evaluate if the proposed certification is able to quote face recognition systems equipped with countermeasures. To this end, we consider a practical use-case scenario for outdoor building access control. Every user must authenticate himself/herself using his/her security badge and face verification is performed to authorize the access to the building. A standard RGB sensor is used for capturing the face and state of the art face recognition software without anti-spoofing countermeasures performs the verification. It is easy to acquire the biometric data to make a fake face by acquiring a video or picture of a real client without his/her complicity using standard commercial cameras. For print and video attacks, we consider the worst case scenario where the impostor hides a high definition camera near the recognition sensor to capture the user during an authentication attempt. The impostor already possesses the necessary biometric data to manufacture video attacks. We evaluate the resistance of protected and unprotected face recognition systems against replay attacks (photo and video) by using the certification methodology previously presented.

5.2.2.1 Use-case1: face verification without anti-spoofing countermeasures.

No anti-spoofing countermeasures are implemented in this first use-case so a very few trials are required to find a fake face capable of bypassing the system as demonstrated in chapter 1. The evaluation of the security level of this system against video attacks are reported in table 5.4.

Elapsed Time: The manufacturing of several fake faces is very easy and fast as printing face pictures takes at most one hour. Printed fake faces or digital ones must be as realistic as possible to have a chance to bypass the system. Multiple trials are conducted on a substitute face recognition system to determine if the manufactured fake faces are working. After experimentation, the impostor manages to bypass the system 8 times over 10 trials. In the end, the manufacturing of a golden fake face takes about 2 days.

To perform the attack, the attacker just need to show the picture/video in front of the sensor

which takes about 3 seconds.

Expertise: The expertise level is layman as no special skills are required to make a golden fake face or to use it.

Knowledge of TOE: The attacker must be aware of the type of sensor used for recognition as well as the acquisition conditions during the authentication process. These information are public as they can be retrieved on site directly with an open access. To perform the attack, only the identification protocol is required which is a public information.

Window of opportunity: The attacker needs to install a hidden camera to get the face data required to manufacture a golden fake face. Therefore an unlimited or easy access is required. In this use-case, the system controls the entrance of a building so people have an open access to it which complies with the requirements to make the attack. When performing the attack, the impostor has to make sure that no one is looking which is quite easy in the considered context.

Equipment: The tools used to implement the full attack are standard: digital camera to retrieve the face data and plain paper with an office printer to make print attacks. A smartphone or iPad are needed for video attacks.

As a result, the rating of the system against photo and video attacks is 10. According to the draft of methodology [CCN11], the rating value is between 0 and 19 so the system fails any level of evaluation. The system is deemed highly vulnerable to video attacks as expected.

5.2.2.2 Use-case2: outdoor building access control with anti-spoofing countermeasures.

The same situation is considered but now motion and texture based anti-spoofing countermeasures are implemented. The evaluation of the security level of this system against video attacks obtains the ratings in table 5.5.

Elapsed Time: The manufacturing of several fake faces is very easy and fast as printing face pictures takes at most one hour. Printed fake faces or digital ones must be as realistic as possible to have a chance to bypass the system. Multiple trials are conducted on a substitute face recognition system to determine if the manufactured fake faces are working. After experimentation, the impostor manages to bypass the system 1 time over 100 trials. In the end, the manufacturing of a golden fake face takes about one month as multiple identities are tested with different video montages. To perform the attack, the attacker just needs to show the picture/video in front of the sensor which takes about 3 seconds.

Expertise, Knowledge of TOE, Window of opportunity, Equipment: Same as before.

The rating in this case is 12, according to the draft of methodology, the rating value is between 0 and 19 so the system fails any level of evaluation. As a result, the face recognition system with

Table 5.4: Security ratings of an unprotected face recognition system.

Criteria	Identification	Exploitation	Total
Elapsed Time			
< = one day	0	2	
< = one week	1	4	3
< = one month	3	6	
< = 3 months	5	8	
< = six months	7	10	
> six months	10	*	
Expertise			
Layman	0	0	0
Proficient	2	0	
Expert	5	0	
Multiples experts	7	0	
Knowledge of Target of Evaluation			
Public	0	0	0
Restricted	2	1	
Sensitive	4	3	
Critical	6	5	
Window of Opportunity			
Unlimited access	0	0	
Easy	1	4	4
Moderate	3	6	
Difficult	5	8	
None	*	*	
Equipment			
None	0	0	
Standard	1	2	3
Specialized	3	4	
Bespoke	5	6	
Multiple Bespoke	7	8	
Total	2	8	**10**

countermeasures is assessed in the same way as the system without countermeasures. The rating granularity does not allow a clear distinction between protected and unprotected systems as the only factor impacted in the evaluation is the time necessary to find a golden fake face.

5.3 Conclusion

This study exhibits the limitations of the proposed methodology. The problem is that both protected and unprotected systems obtain similar ratings. The success rate of an attack largely decreases when anti-spoofing countermeasures are implemented and more trials are required to manage a successful spoofing attempt. The anti-spoofing performance is somehow taken into account in the Elapsed time criterion to reflect the difficulty to determine the 'golden fake face' but its impact on the final score is negligible. Besides, multiple trials require the manufacturing of several fake faces which becomes costly as the number of trials rises but this factor is overlooked in the evaluation. To take into account these aspects in the certification process, we propose two measures in terms of time and cost. Given the success rate of an attack SR, the average time required to find a golden fake face is given by $t = t_0 * 1/SR$ where t_0 is the time to manufacture one fake face.

Table 5.5: Security ratings of a protected face recognition system.

Criteria	Identification	Exploitation	Total
Elapsed Time			
< = one day	0	2	
< = one week	1	4	3
< = one month	3	6	
< = 3 months	5	8	
< = six months	7	10	
> six months	10	*	
Expertise			
Layman	0	0	0
Proficient	2	0	
Expert	5	0	
Multiples experts	7	0	
Knowledge of Target of Evaluation			
Public	0	0	0
Restricted	2	1	
Sensitive	4	3	
Critical	6	5	
Window of Opportunity			
Unlimited access	0	0	
Easy	1	4	4
Moderate	3	6	
Difficult	5	8	
None	*	*	
Equipment			
None	0	0	
Standard	1	2	3
Specialized	3	4	
Bespoke	5	6	
Multiple Bespoke	7	8	
Total	4	8	12

The associated cost is $c = c_0 * 1/SR$ where c_0 corresponds to the cost related to the fabrication of one fake face. A special rating of both measures is needed so that the final rating reflects properly the resistance of a face verification system against spoofing attacks.

Furthermore, the difficulty to acquire the biometric data to manufacture a realistic fake face is overlooked. We assumed that the acquisition of the face negative was possible via a hidden HD camera. However, this scenario is unlikely to happen in high security use-cases and the most realistic way is to find videos or pictures of a valid user on the internet. Some image processing may be needed to transform the image into a successful attack. The difficulty to find a successful attack using internet pictures/videos has to be taken into account into the ratings.

Conclusion and perspectives

Conclusions

The development of public spoofing databases has boosted significantly the activity on face anti-spoofing research and a large panel of countermeasures have been developed since 2010. This doctoral dissertation is motivated by the evaluation of these protection measures for certification. A large part of this work is about assessing the vulnerabilities of 2D face biometric systems against spoofing attacks. Taking advantage of recent release of public databases, we focussed our study on software-based anti-spoofing methods implemented on standard face identification systems relying on RGB video acquisitions of the face.

The development of new spoofing attacks and new anti-spoofing countermeasures are closely related and both aspects have been explored throughout this work with a complete review of state of the art in fake face forgery and anti-spoofing countermeasure design. The most recent spoofing strategies have been identified and a special effort has been brought to redefine a suitable taxonomy of software-based countermeasures highlighting the type of discriminant cues they rely upon. Although evaluation on mask attacks are missing in this study, the line of work followed for the development of new countermeasures anticipates these attacks and concentrates on the face region only.

We proposed three different approaches to tackle the anti-spoofing problem. In the second chapter, data-driven methods based on texture are investigated. As texture information is highly dependent on acquisition conditions, a complete study of the well-known LBP countermeasure is conducted to derive a unified framework to deal with multiple sensors and multiple databases. We have demonstrated that the face region is of great importance when exploiting texture cues as well as is the face geometric normalization. Especially, it appears that normalizing faces with a 54 pixels interocular distance is sufficient to extract discriminant texture information regardless of the sensor resolution. Also, better results are obtained by averaging features over time. Under this framework, state of the art texture-based methods are re-evaluated along with two new variants of the traditional LBP descriptor, namely CLBP and HSI-LBP. Taking advantage of contrast and color information, both approaches yield improvements over the traditional LBP countermeasure. In particular, the proposed HSI-LBP feature demonstrates state of the art results on the ReplayAttack, CASIA and MSU databases with respect to texture methods focusing on only the face region only.

In the third chapter, a novel motion-based countermeasure based on rigid and non-rigid face movements is proposed. Using face tracking CLNF framework, face motion is captured into a set of 5 rigid parameters and 34 non-rigid parameters. Discriminant features are computed using the Fisher framework by encoding short motion sequences into a mid-level representation. The proposed method demonstrates very good photo attack detection even when limited motion is imposed by the authentication protocol and it allows real time detection. Video attacks are well detected provided that sufficient hand-shaking motion is present when presenting the fake face in front of the sensor. Hence, competitive results are obtained on ReplayAttack and CASIA databases compared to state of the art methods using optical flow. Besides, the proposed method is robust to camera shake as verified by experiments on MSU database.

In the fourth chapter, a model-based approach based on the recapturing process involved when

performing a spoofing attack is proposed. The proposed recapturing model considers radiometric and blur distortions to explain the disparity between real and fake faces. Using enrolment samples, both distortions are estimated separately and recovered model parameters are used for classification. Radiometric distortions are modelled by color scaling, color offset and gamma non-linearity. Under known illumination conditions, the radiometric features are able to detect almost perfectly fake faces on ReplayAttack, CASIA (H protocol) and MSU databases. In parallel, the blur kernel is estimated by a blind deconvolution technique. Unfortunately, the selected debluring method is not able to estimate precisely the blur distortions but decent detection performance is achieved for high quality acquisitions. Also, a short investigation about the synthesis of spoofing attacks for new identities have obtained promising results in predicting color distortions.

Finally, the last chapter shows the limitations of the certification methodology originally developed for fingerprint technology as unprotected and protected face biometrics systems obtain almost the same security ratings. Ratings need to be adapted and we proposed a few changes to better reflect the resistance of systems against face spoofing attacks.

Future works

Fake face detection is a difficult problem as it suffers from many sources of variability:

- large diversity of face profiles
- various attack types and attack scenarios
- pose and facial expression variations
- illumination variations
- image quality

To tackle these problems, multiple databases have been considered. In the course of this thesis, we have supposed that acquisition conditions are consistent between training and testing the anti-spoofing countermeasures. However, for certain types of applications such as face identification on laptops or mobile, these assumptions are no longer verified and further development is required in that regard and cross-database experiments must be considered.

Additionally, experiments on mask attacks have not been considered despite having access to the public 3D mask attack database or the MORPHO mask attack database (provided by Safran) because the format of the recordings are inconsistent with the color video acquisitions required to extract the proposed texture and motion features. Mask spoofing datasets are rather designed for evaluations of 3D face recognition systems and acquisitions of realistic masks with decent quality RGB sensors are lacking for further development of 2D face anti-spoofing countermeasures.

This work has open diverse interesting research directions for further developments. We list below a few ideas that are considered in perspective:

- In chapter 2, improvements of LBP countermeasures are proposed based on color and contrast information. Combining the multi-scale LBP computation strategy, color and contrast information simultaneously is the next step.

- In chapter 3, rigid and non-rigid motion sequences are transformed into discriminant mid-level features using the Fisher framework. Another approach is to consider the motion signatures as time series and to extract discriminant information using signal processing tools from this domain.

- In chapter 4, the estimation of radiometric distortions is done by the simpler per-channel Gamma model as unstable parameter estimation is obtained when considering the coupled Gamma model. Further constraints are required to regularize the solution for practical use of the more flexible coupled Gamma model. Also, the proposed blur kernel estimation is not good enough to capture motion blur and anti-aliasing blur generated by the recapturing process. Faces have a limited amount of edges to guide the estimation, another blur metric should be used instead. One interesting research direction is to use edge profiles obtained from the face structure used in the face tracking phase (68 facial landmarks) and to estimate the blur magnitude for each edge profile. Another focal point of future work is the synthesis of spoofing attacks for new identities. We believe that semi-coupled dictionary learning techniques have potential for synthesizing the texture components of spoofing attacks. The synthesis is not limited to face images and can be extended to features. The final goal is to perform a person-specific fake face detection to better handle the diversity of face profiles.

- In chapter 5, the current certification methodology is not satisfactory as no anti-spoofing countermeasures validate the security ratings. Strong connections with the BEAT European project is evident and a collaboration is to look out for in the future.

- Additionally, the fusion of the proposed methods must be considered to exploit fully all the complementary cues presented in this work. As discriminant cues vary from one attack type to the other, an ensemble of classifiers is potentially a good way to integrate multiple cues for fake face detection.

Bibliography

[Anjos11] André Anjos and Sébastien Marcel. Counter-measures to photo attacks in face recognition: A public database and a baseline. In *Biometrics (IJCB), 2011 International Joint Conference On*, pages 1–7. 2011.

[Anjos14] A. Anjos, M.M. Chakka, and S. Marcel. Motion-based counter-measures to photo attacks in face recognition. *IET Biometrics*, 3(3):147–158, September 2014.

[Bai10] Jiamin Bai, Tian-Tsong Ng, Xinting Gao, and Yun-Qing Shi. Is physics-based liveness detection truly possible with a single image? In *Circuits and Systems (ISCAS), Proceedings of 2010 IEEE International Symposium On*, pages 3425–3428. 2010.

[Baltru16] Tadas Baltru, Peter Robinson, Louis-Philippe Morency, and others. OpenFace: An open source facial behavior analysis toolkit. In *2016 IEEE Winter Conference on Applications of Computer Vision (WACV)*, pages 1–10. IEEE, 2016.

[Baltrusaitis13] Tadas Baltrusaitis, Peter Robinson, and Louis-Philippe Morency. Constrained local neural fields for robust facial landmark detection in the wild. In *Proceedings of the IEEE International Conference on Computer Vision Workshops*, pages 354–361. 2013.

[Bao09] Wei Bao, Hong Li, Nan Li, and Wei Jiang. A liveness detection method for face recognition based on optical flow field. In *Image Analysis and Signal Processing, 2009. IASP 2009. International Conference On*, pages 233–236. 2009.

[Benlamoudi15] Azeddine Benlamoudi, Djamel Samai, Abdelkrim Ouafi, Salah Eddine Bekhouche, Abdelmalik Taleb-Ahmed, and Abdenour Hadid. Face spoofing detection using local binary patterns and Fisher Score. In *2015 3rd International Conference on Control, Engineering Information Technology (CEIT)*, pages 1–5. May 2015.

[Bharadwaj13] Samarth Bharadwaj, Tejas I. Dhamecha, Mayank Vatsa, and Richa Singh. Computationally Efficient Face Spoofing Detection with Motion Magnification. In *Computer Vision and Pattern Recognition Workshops (CVPRW), 2013 IEEE Conference On*, pages 105–110. 2013.

[Bharadwaj14] Samarth Bharadwaj, Tejas I. Dhamecha, Mayank Vatsa, and Richa Singh. Face anti-spoofing via motion magnification and multifeature videolet aggregation. 2014.

[Bianconi07] Francesco Bianconi and Antonio Fernández. Evaluation of the effects of Gabor filter parameters on texture classification. *Pattern Recognition*, 40(12):3325–3335, December 2007.

[Castrillón07] M. Castrillón, O. Déniz, C. Guerra, and M. Hernández. ENCARA2: Real-time detection of multiple faces at different resolutions in video streams. *Journal of Visual Communication and Image Representation*, 18(2):130–140, April 2007.

[CCN11] CCN. Characterizing Attacks to Fingerprint Verification Mechanisms. Technical report, National Cryptologic Centre, 2011.

[Chakka11] Murali Mohan Chakka, Andre Anjos, Sebastien Marcel, Roberto Tronci, Daniele Muntoni, Gianluca Fadda, Maurizio Pili, Nicola Sirena, Gabriele Murgia, and Marco Ristori. Competition on counter measures to 2-d facial spoofing attacks. In *Biometrics (IJCB), 2011 International Joint Conference On*, pages 1–6. 2011.

[Chakrabarti09] Ayan Chakrabarti, Daniel Scharstein, and Todd Zickler. An Empirical Camera Model for Internet Color Vision. In *BMVC*, volume 1, page 4. Citeseer, 2009.

[Chang11] Chih-Chung Chang and Chih-Jen Lin. LIBSVM: A library for support vector machines. *ACM Transactions on Intelligent Systems and Technology (TIST)*, 2(3):27, 2011.

[Chatzichristofis08a] Savvas A. Chatzichristofis and Yiannis S. Boutalis. CEDD: Color and edge directivity descriptor: A compact descriptor for image indexing and retrieval. In *Computer Vision Systems*, pages 312–322. Springer, 2008.

[Chatzichristofis08b] Savvas A. Chatzichristofis and Yiannis S. Boutalis. FCTH: Fuzzy Color and Texture Histogram - A Low Level Feature for Accurate Image Retrieval. pages 191–196. IEEE, 2008.

[Chaudhry09] Rizwan Chaudhry, Avinash Ravichandran, Gregory Hager, and René Vidal. Histograms of oriented optical flow and binet-cauchy kernels on nonlinear dynamical systems for the recognition of human actions. In *Computer Vision and Pattern Recognition, 2009. CVPR 2009. IEEE Conference On*, pages 1932–1939. 2009.

[Chingovska12] Ivana Chingovska, André Anjos, and Sébastien Marcel. On the effectiveness of local binary patterns in face anti-spoofing. In *Biometrics Special Interest Group (BIOSIG), 2012 BIOSIG-Proceedings of the International Conference of The*, pages 1–7. IEEE, 2012.

[Chingovska13a] Ivana Chingovska. The 2nd Competition on Counter Measures to 2D Face Spoofing Attacks. In *Biometrics (ICB), 2013 International Conference*. 2013.

[Chingovska13b] Ivana Chingovska, André Anjos, and Sébastien Marcel. Anti-spoofing in action: Joint operation with a verification system. *Computer Vision and Pattern Recognition Workshops (CVPRW), 2013 IEEE Conference*, 2013.

[Chingovska15] I. Chingovska and A. Rabello dos Anjos. On the Use of Client Identity Information for Face Antispoofing. *IEEE Transactions on Information Forensics and Security*, 10(4):787–796, April 2015.

[Clausi02] David A. Clausi. An analysis of co-occurrence texture statistics as a function of grey level quantization. *Canadian Journal of remote sensing*, 28(1):45–62, 2002.

[Crete07] Frederique Crete, Thierry Dolmiere, Patricia Ladret, and Marina Nicolas. The blur effect: Perception and estimation with a new no-reference perceptual blur metric. In *Electronic Imaging 2007*, pages 64920I–64920I. International Society for Optics and Photonics, 2007.

[De Marsico12] Maria De Marsico, Michele Nappi, Daniel Riccio, and J. Dugelay. Moving face spoofing detection via 3D projective invariants. In *Biometrics (ICB), 2012 5th IAPR International Conference On*, pages 73–78. 2012.

[Duc09] Nguyen Minh Duc and Bui Quang Minh. Your face is NOT your password Face Authentication ByPassing Lenovo–Asus–Toshiba. *Black Hat Briefings*, 2009.

[Eid11] A H Eid, M N Ahmed, B E Cooper, and E E Rippetoe. Characterization of Electrophotographic Print Artifacts: Banding, Jitter, and Ghosting. *IEEE Transactions on Image Processing*, 20(5):1313–1326, May 2011.

[Erdogmus13] Nesli Erdogmus and Sebastien Marcel. Spoofing 2D face recognition systems with 3D masks. In *Biometrics Special Interest Group (BIOSIG), 2013 International Conference of The*, pages 1–8. 2013.

[Freitas Pereira13] Tiago Freitas Pereira, André Anjos, José Mario De Martino, and Sébastien Marcel. Can face anti-spoofing countermeasures work in a real world scenario? *ICB2013*, 2013.

[Galbally10] Javier Galbally, Chris McCool, Julian Fierrez, Sebastien Marcel, and Javier Ortega-Garcia. On the vulnerability of face verification systems to hill-climbing attacks. *Pattern Recognition*, 43(3):1027–1038, 2010.

[Galbally14a] Javier Galbally and Sébastien Marcel. Face Anti-Spoofing Based on General Image Quality Assessment. 2014.

[Galbally14b] Javier Galbally, Sebastien Marcel, and Julian Fierrez. Biometric Antispoofing Methods: A Survey in Face Recognition. *IEEE Access*, 2:1530–1552, 2014.

[Galbally14c] Javier Galbally, Sebastien Marcel, and Julian Fierrez. Image Quality Assessment for Fake Biometric Detection: Application to Iris, Fingerprint, and Face Recognition. *IEEE Transactions on Image Processing*, 23(2):710–724, February 2014.

[Gao10] Xinting Gao, Tian-Tsong Ng, Bo Qiu, and Shih-Fu Chang. Single-view recaptured image detection based on physics-based features. In *2010 IEEE International Conference on Multimedia and Expo (ICME)*, pages 1469–1474. July 2010.

[Gomez-Barrero12] Marta Gomez-Barrero, Javier Galbally, Julian Fierrez, and Javier Ortega-Garcia. Face verification put to test: A hill-climbing attack based on the uphill-simplex algorithm. In *Biometrics (ICB), 2012 5th IAPR International Conference On*, pages 40–45. 2012.

[Gorodnichy14] D. Gorodnichy, Eric Granger, Paolo Radtke, Contract Scientific Authority, and Pierre Meunier. Survey of commercial technologies for face recognition in video. *CBSA, Border Technology Division, Tech. Rep*, 22, 2014.

[Gross03] Ralph Gross and Vladimir Brajovic. An image preprocessing algorithm for illumination invariant face recognition. In *Audio-and Video-Based Biometric Person Authentication*, pages 10–18. Springer, 2003.

[Guo10a] Zhenhua Guo and David Zhang. A completed modeling of local binary pattern operator for texture classification. *Image Processing, IEEE Transactions on*, 19(6):1657–1663, 2010.

[Guo10b] Zhenhua Guo, Lei Zhang, and David Zhang. Rotation invariant texture classification using LBP variance (LBPV) with global matching. *Pattern Recognition*, 43(3):706–719, March 2010.

[Haralick73] Robert M. Haralick, Karthikeyan Shanmugam, and Its' Hak Dinstein. Textural features for image classification. *Systems, Man and Cybernetics, IEEE Transactions on*, (6):610–621, 1973.

[He13] Kaiming He, Jian Sun, and Xiaoou Tang. Guided Image Filtering. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 35(6):1397–1409, June 2013.

[Hiscocks11] Peter Hiscocks. Measuring luminance with a digital camera. 2011.

[Hsu03] Chih-Wei Hsu, Chih-Chung Chang, and Chih-Jen Lin. *A Practical Guide to Support Vector Classification*. 2003.

[Jain08] Anil K Jain, Karthik Nandakumar, and Abhishek Nagar. Biometric Template Security. *EURASIP Journal on Advances in Signal Processing*, 2008(1):579416, 2008.

[Joshi10] Neel Joshi, Wojciech Matusik, Edward H. Adelson, and David J. Kriegman. Personal photo enhancement using example images. *ACM Transactions on Graphics*, 29(2):1–15, March 2010.

[Kim12] Gahyun Kim, Sungmin Eum, Jae Kyu Suhr, Dong Ik Kim, Kang Ryoung Park, and Jaihie Kim. Face liveness detection based on texture and frequency analyses. In *Biometrics (ICB), 2012 5th IAPR International Conference On*, pages 67–72. IEEE, 2012.

[Kimmel03] Ron Kimmel, Michael Elad, Doron Shaked, Renato Keshet, and Irwin Sobel. A variational framework for retinex. *International Journal of Computer Vision*, 52(1):7–23, 2003.

[Kobayashi12] Takumi Kobayashi and Nobuyuki Otsu. Motion recognition using local autocorrelation of space–time gradients. *Pattern Recognition Letters*, 33(9):1188–1195, July 2012.

[Kollreider07a] K. Kollreider, H. Fronthaler, and J. Bigun. Non-intrusive liveness detection by face images. *Image and Vision Computing*, 27(3):233–244, 2007.

[Kollreider07b] Klaus Kollreider, Hartwig Fronthaler, Maycel Isaac Faraj, and Josef Bigun. Real-Time Face Detection and Motion Analysis With Application in Liveness Assessment. *IEEE Transactions on Information Forensics and Security*, 2(3):548–558, 2007.

[Kollreider08] Klaus Kollreider, Hartwig Fronthaler, and Josef Bigun. Verifying liveness by multiple experts in face biometrics. In *Computer Vision and Pattern Recognition Workshops, 2008. CVPRW'08. IEEE Computer Society Conference On*, pages 1–6. 2008.

[Komulainen13] Jukka Komulainen, Abdenour Hadid, and Matti Pietikäinen. Face spoofing detection using dynamic texture. In *Computer Vision-ACCV 2012 Workshops*, pages 146–157. 2013.

[Kose12] Neslihan Kose and Jean-Luc Dugelay. Classification of captured and recaptured images to detect photograph spoofing. In *Informatics, Electronics & Vision (ICIEV), 2012 International Conference On*, pages 1027–1032. 2012.

[Kose13a] Neslihan Kose and Jean-Luc Dugelay. Countermeasure for the Protection of Face Recognition Systems Against Mask Attacks. In *FG 2013, 10th IEEE International Conference on Automatic Face and Gesture Recognition, 22-26 April 2013, Shanghai, China*. 2013.

[Kose13b] Neslihan Kose and Jean-Luc Dugelay. Reflectance analysis based countermeasure technique to detect face mask attacks. In *Digital Signal Processing (DSP), 2013 18th International Conference On*, pages 1–6. 2013.

[Kose13c] Neslihan Kose and Jean-Luc Dugelay. Shape and Texture Based Countermeasure to Protect Face Recognition Systems Against Mask Attacks. In *Computer Vision and Pattern Recognition Workshops (CVPRW), 2013 IEEE Conference On*, pages 111–116. 2013.

[Lee05] Kuang-Chih Lee, Jeffrey Ho, Ming-Hsuan Yang, and David Kriegman. Visual tracking and recognition using probabilistic appearance manifolds. *Computer Vision and Image Understanding*, 99(3):303–331, September 2005.

[Li04] Jiangwei Li, Yunhong Wang, Tieniu Tan, and Anil K. Jain. Live face detection based on the analysis of fourier spectra. In *Defense and Security*, pages 296–303. 2004.

[Luka06] J. Luka, J. Fridrich, and M. Goljan. Digital Camera Identification From Sensor Pattern Noise. *IEEE Transactions on Information Forensics and Security*, 1(2):205–214, June 2006.

[Maatta11] Jukka Maatta, Abdenour Hadid, and Matti Pietikainen. Face spoofing detection from single images using micro-texture analysis. In *Biometrics (IJCB), 2011 International Joint Conference On*, pages 1–7. 2011.

[Mairal09] Julien Mairal, Francis Bach, Jean Ponce, and Guillermo Sapiro. Online dictionary learning for sparse coding. In *Proceedings of the 26th Annual International Conference on Machine Learning*, pages 689–696. ACM, 2009.

[Marcel14] Sébastien Marcel, Mark S. Nixon, and Stan Z. Li. *Handbook of Biometric Anti-Spoofing: Trusted Biometrics Under Spoofing Attacks*. Springer Publishing Company, Incorporated, 2014.

[Martin97] Alvin Martin, George Doddington, Terri Kamm, Mark Ordowski, and Mark Przybocki. The DET curve in assessment of detection task performance. Technical report, DTIC Document, 1997.

[Marziliano02] Pina Marziliano, Frederic Dufaux, Stefan Winkler, and Touradj Ebrahimi. A no-reference perceptual blur metric. In *Image Processing. 2002. Proceedings. 2002 International Conference On*, volume 3, pages III–57. IEEE, 2002.

[Ojala02] Timo Ojala, Matti Pietikainen, and Topi Maenpaa. Multiresolution gray-scale and rotation invariant texture classification with local binary patterns. *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, 24(7):971–987, 2002.

[P. Ekman78] P. Ekman and W. Friesen. Facial Action Coding System: A Technique for the Measurement of Facial Movement. *Consulting Psychologists Press*, 1978.

[Pan07] Gang Pan, Lin Sun, Zhaohui Wu, and Shihong Lao. Eyeblink-based anti-spoofing in face recognition from a generic webcamera. In *Computer Vision, 2007. ICCV 2007. IEEE 11th International Conference On*, pages 1–8. 2007.

[Pan14] Jinshan Pan, Zhe Hu, Zhixun Su, and Ming-Hsuan Yang. Deblurring face images with exemplars. In *Computer Vision–ECCV 2014*, pages 47–62. Springer, 2014.

[Parmar14] Divyarajsinh N. Parmar and Brijesh B. Mehta. Face Recognition Methods & Applications. *arXiv preprint arXiv:1403.0485*, 2014.

[Peixoto11] Bruno Peixoto, Carolina Michelassi, and Anderson Rocha. Face liveness detection under bad illumination conditions. In *Image Processing (ICIP), 2011 18th IEEE International Conference On*, pages 3557–3560. 2011.

[Perronnin07] Florent Perronnin and Christopher Dance. Fisher kernels on visual vocabularies for image categorization. In *2007 IEEE Conference on Computer Vision and Pattern Recognition*, pages 1–8. IEEE, 2007.

[Perronnin10] Florent Perronnin, Jorge Sánchez, and Thomas Mensink. Improving the fisher kernel for large-scale image classification. In *European Conference on Computer Vision*, pages 143–156. Springer, 2010.

[Pietikäinen11] Matti Pietikäinen, Abdenour Hadid, Guoying Zhao, and Timo Ahonen. Local Binary Patterns for Still Images. In *Computer Vision Using Local Binary Patterns*, volume 40, pages 13–47. Springer London, London, 2011.

[Pinto12] Allan da Silva Pinto, Helio Pedrini, William Schwartz, and Anderson Rocha. Video-Based Face Spoofing Detection through Visual Rhythm Analysis. In *Graphics, Patterns and Images (SIBGRAPI), 2012 25th SIBGRAPI Conference On*, pages 221–228. 2012.

[Pinto15a] A. Pinto, H. Pedrini, W. Schwartz, and A. Rocha. Face Spoofing Detection Through Visual Codebooks of Spectral Temporal Cubes. *IEEE Transactions on Image Processing*, PP(99):1–1, 2015.

[Pinto15b] A. Pinto, W. Robson Schwartz, H. Pedrini, and A. De Rezende Rocha. Using Visual Rhythms for Detecting Video-Based Facial Spoof Attacks. *IEEE Transactions on Information Forensics and Security*, 10(5):1025–1038, May 2015.

[Pitie05] Francois Pitie, Anil C. Kokaram, and Rozenn Dahyot. N-dimensional probability density function transfer and its application to color transfer. In *Computer Vision, 2005. ICCV 2005. Tenth IEEE International Conference On*, volume 2, pages 1434–1439. IEEE, 2005.

[Ratha01a] Nalini K. Ratha, Jonathan H. Connell, and Ruud M. Bolle. An analysis of minutiae matching strength. In *Audio-and Video-Based Biometric Person Authentication*, pages 223–228. 2001.

[Ratha01b] Nalini K. Ratha, Jonathan H. Connell, and Ruud M. Bolle. Enhancing security and privacy in biometrics-based authentication systems. *IBM systems journal*, 40(3):614–634, 2001.

[Saragih10] Jason M. Saragih, Simon Lucey, and Jeffrey F. Cohn. Deformable Model Fitting by Regularized Landmark Mean-Shift. *International Journal of Computer Vision*, 91(2):200–215, September 2010.

[Schwartz11] W. Schwartz, Anderson Rocha, and Helio Pedrini. Face spoofing detection through partial least squares and low-level descriptors. In *Biometrics (IJCB), 2011 International Joint Conference On*, pages 1–8. 2011.

[Soh99] L. K. Soh and C. Tsatsoulis. Texture analysis of SAR sea ice imagery using gray level co-occurrence matrices. *IEEE Transactions on Geoscience and Remote Sensing*, 37(2):780–795, March 1999.

[Son Vu10] Ngoc Son Vu. *Contributions à La Reconnaissance de Visages à Partir d'une Seule Image et Dans Un Contexte Non-Contrôlé*. Ph.D. thesis, Université de Grenoble et Institut Polytechnique de Grenoble, 2010.

[Soutar02] Colin Soutar. Biometric system security. *White Paper, Bioscrypt, http://www. bioscrypt. com*, 2002.

[Štruc10] Vitomir Štruc and Nikola Pavešić. The Complete Gabor-Fisher Classifier for Robust Face Recognition. *EURASIP Journal on Advances in Signal Processing*, 2010(1):847680, 2010.

[Tan05] Robby T. Tan and Katsushi Ikeuchi. Separating reflection components of textured surfaces using a single image. *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, 27(2):178–193, 2005.

[Tan10a] Xiaoyang Tan, Yi Li, Jun Liu, and Lin Jiang. Face liveness detection from a single image with sparse low rank bilinear discriminative model. In *Computer Vision–ECCV 2010*, pages 504–517. Springer, 2010.

[Tan10b] Xiaoyang Tan and Bill Triggs. Enhanced Local Texture Feature Sets for Face Recognition Under Difficult Lighting Conditions. *IEEE Transactions on Image Processing*, 19(6):1635–1650, June 2010.

[Tirunagari15] S. Tirunagari, N. Poh, D. Windridge, A. Iorliam, N. Suki, and A.T.S. Ho. Detection of Face Spoofing Using Visual Dynamics. *IEEE Transactions on Information Forensics and Security*, 10(4):762–777, April 2015.

[Trefny10] Jirí Trefny and Jirí Matas. Extended set of local binary patterns for rapid object detection. In *Proceedings of the Computer Vision Winter Workshop*, volume 2010. 2010.

[Tronci11] Roberto Tronci, Daniele Muntoni, Gianluca Fadda, Maurizio Pili, Nicola Sirena, Gabriele Murgia, Marco Ristori, and Fabio Roli. Fusion of multiple clues for photo-attack detection in face recognition systems. In *Biometrics (IJCB), 2011 International Joint Conference On*, pages 1–6. 2011.

[Wang13] Tao Wang, Jianwei Yang, Zhen Lei, Shengcai Liao, and Stan Z. Li. Face Liveness Detection Using 3D Structure Recovered from a Single Camera. 2013.

[Waris13] Muhammad-Adeel Waris, Honglei Zhang, Iftikhar Ahmad, Serkan Kiranyaz, and Moncef Gabbouj. Analysis of textural features for face biometric anti-spoofing. In *21st European Signal Processing Conference (EUSIPCO 2013)*, pages 1–5. IEEE, 2013.

[Waris14] Muhammad-Adeel Waris. Securing face biometric systems from spoofing attacks. Technical report, TAMPERE UNIVERSITY OF TECHNOLOGY, 2014.

[Wen15] Di Wen, Hu Han, and A.K. Jain. Face Spoof Detection With Image Distortion Analysis. *IEEE Transactions on Information Forensics and Security*, 10(4):746–761, April 2015.

[Wu12] Hao-Yu Wu, Michael Rubinstein, Eugene Shih, John Guttag, Frédo Durand, and William Freeman. Eulerian video magnification for revealing subtle changes in the world. *ACM Transactions on Graphics (TOG)*, 31(4):65, 2012.

[Yan12] Junjie Yan, Zhiwei Zhang, Zhen Lei, Dong Yi, and Stan Z. Li. Face liveness detection by exploring multiple scenic clues. In *Control Automation Robotics & Vision (ICARCV), 2012 12th International Conference On*, pages 188–193. 2012.

[Yang13] Jianwei Yang, Zhen Lei, Shengcai Liao, and Stan Z. Li. Face Liveness Detection with Component Dependent Descriptor. *Biometrics(ICB), 2013 International Conference*, 2013.

[Yang15] Jianwei Yang, Zhen Lei, Dong Yi, and S.Z. Li. Person-Specific Face Antispoofing With Subject Domain Adaptation. *IEEE Transactions on Information Forensics and Security*, 10(4):797–809, April 2015.

[Yu08] Hang Yu, Tian-Tsong Ng, and Qibin Sun. Recaptured photo detection using specularity distribution. In *Image Processing, 2008. ICIP 2008. 15th IEEE International Conference On*, pages 3140–3143. IEEE, 2008.

[Zhang12] Zhiwei Zhang, Junjie Yan, Sifei Liu, Zhen Lei, Dong Yi, and Stan Z. Li. A face antispoofing database with diverse attacks. In *Biometrics (ICB), 2012 5th IAPR International Conference On*, pages 26–31. 2012.

[Zhu10] Chao Zhu, Charles-Edmond Bichot, and Liming Chen. Multi-scale Color Local Binary Patterns for Visual Object Classes Recognition. pages 3065–3068. IEEE, August 2010.

Résumé — Les systèmes d'identification faciale sont en plein essor et se retrouvent de plus en plus dans des produits grand public tels que les smartphones et les ordinateurs portables. Cependant, ces systèmes peuvent être facilement bernés par la présentation par exemple d'une photo imprimée de la personne ayant les droits d'accès au système. Cette thèse s'inscrit dans le cadre du projet ANR BIOFENCE qui vise à développer une certification des systèmes biométriques veine, iris et visage permettant aux industriels de faire valoir leurs innovations en termes de protection. L'objectif de cette thèse est double, d'abord il s'agit de développer des mesures de protection des systèmes 2D d'identification faciale vis à vis des attaques connues à ce jour (photos imprimées, photos ou vidéos sur un écran, masques) puis de les confronter à la méthodologie de certification développée au sein du projet ANR. Dans un premier temps, un état de l'art général des attaques et des contremesures est présenté en mettant en avant les méthodes algorithmiques (« software ») par rapport aux méthodes hardware. Ensuite, plusieurs axes sont approfondis au cours de ce travail. Le premier concerne le développement d'une contremesure basée sur une analyse de texture et le second concerne le développement d'une contre-mesure basée sur une analyse de mouvement. Ensuite, une modélisation du processus de recapture pour différencier un faux visage d'un vrai est proposée. Une nouvelle méthode de protection est développée sur ce concept en utilisant les données d'enrolment des utilisateurs et un premier pas est franchi dans la synthèse d'attaque pour un nouvel utilisateur à partir de sa donnée d'enrolment. Enfin, la méthodologie de certification développée pour les systèmes à empreintes digitales est évaluée pour les systèmes d'identification facial.

Mots clés : Biométrie visage, contre-mesures, faux visages, analyse du mouvement, analyse de texture.

Abstract — Face identification systems are growing rapidly and invade the consumer market with security products in smartphones, computers and banking. However, these systems are easily fooled by presenting a picture of the person having legitimate access to the system. This thesis is part of the BIOFENCE project which aim to develop a certification of biometric systems in order for industrials to promote their innovations in terms of protection. Our goal is to develop new anti-spoofing countermeasures for 2D face biometric systems and to evaluate the certification methodology on protected systems. First, a general state of the art in face spoofing attack forgery and in anti-spoofing protection measures is presented. Then texture-based countermeasures and motion-based countermeasures are investigated leading to the development of two novel countermeasures. Then, the recapturing process is modelled and a new fake face detection approach is proposed based on this model. Taking advantage of enrolment samples from valid users, a first step toward the synthesis of spoofing attacks for new users is taken. Finally, the certification methodology originally developed for fingerprint technology is evaluated on face biometric systems.

Keywords: Biometrics, face anti-spoofing, fake faces, motion analysis, texture analysis.

Gipsa-lab, 11 Rue des Mathématiques
38400 Saint-Martin-d'Hères